Genetic Variation in the Wood Chemistry of *Eucalyptus globulus*

Fiona S. Poke BSc (Hons)

Submitted in fulfilment of the requirements for the Degree of Doctor of Philosophy

University of Tasmania

April, 2006

DECLARATION

This thesis contains no material that has been accepted for a degree or diploma by any tertiary institution. To the best of my knowledge and belief, this thesis contains no material previously published or written by another person except where due acknowledgment is made in the text.

This thesis may be made available for loan and limited copying in accordance with the Copyright Act 1968.

Fiona S. Poke

ACKNOWLEDGMENTS

Many thanks must go to all the people who have helped me during the last three and a half years. To my supervisors Jim Reid, Carolyn Raymond, René Vaillancourt and Bruce Greaves, I could not have asked for better guidance. Thanks to the co-authors on my published manuscripts: Brad Potts for his input into Chapter 1 and help with executing the statistical analyses in Chapter 5; Dorothy Steane for her instruction on the phylogenetic analysis in Chapter 7; Darren Martin for his assistance with the recombination detection analysis in Chapter 7; and Judith Wright for the collection of NIR spectra and development of the calibrations in Chapters 2 and 3. I would also like to thank Linda Ballard, Leon Savage and Leigh Johnson for their technical assistance, plus Greg Jordan for his help with the statistical analyses in Chapter 6.

A very special thank you goes to my colleagues in the School of Plant Science for making my PhD years so enjoyable. To those I have been sharing an office with, eating lunch with and having a "knock-off" drink with, who have been there for the ride – Corinne, Prue, Beck, Matt and Brett, you guys are the best.

Most of all I would like to thank those people who have supported me the most, my friends and family. To my parents, Lance and Kaye, and my brother Rohan thanks for everything. To Simon, thanks for your understanding, support and encouragement, and most of all, thanks for keeping life fun. And lastly, to my wonderful, amazing friends Jacs, Rach, Brad, Justin, Nige, Bam, Deb, Anna, Simon who define my fun times and who will help me celebrate the end of this era!

CONTENTS

ABSTRACT

Eucalyptus globulus is grown in temperate regions throughout the world predominantly for the production of kraft pulp. Kraft pulping involves chemically removing most of the lignin and extractives and some of the hemicellulose from the cellulose fibres. The amount of lignin and extractives in the wood is therefore important. The aim of this study was to strengthen knowledge of genetic variation in the chemical wood properties of *E. globulus*. Methods were developed for fast and simple assessment of these traits. These will benefit quantitative and molecular approaches to breeding, both of which were explored in this thesis. In addition, the utility of a lignin biosynthesis gene for phylogenetic analysis in *Eucalyptus* was investigated.

Measuring wood chemistry using traditional chemical methods is costly and time-consuming. Near infrared reflectance (NIR) analysis was explored as an alternative to these for the prediction of extractives, lignin and cellulose contents from both ground and solid wood samples. Good calibrations were developed for each of these traits with high correlation coefficients (R^2 of 0.62 to 0.93) and standard errors of less than 1.37%. All calibrations were validated using a separate set of samples with strong correlations obtained between predicted and laboratory values (R^2 of 0.67 to 0.99), with the exception of acid-soluble lignin content for solid wood (R^2 of 0.12). NIR analysis was found to be a reliable predictor of the wood chemistry of *E. globulus,* with solid wood shown to be a good alternative to ground wood samples.

Solid wood NIR calibrations were then used to assess within-tree variation in the wood chemistry of *E. globulus*. Wood chemistry was found to vary within-tree, with extractives content significantly decreasing with height in the tree, but not varying radially. Bark-to-pith variation was observed for cellulose and lignin contents, the former decreasing and the latter increasing. However, for cellulose content this was generally not significant, while for lignin content significance was found between 20 and 60% of tree height. These traits showed no significant height variation. Results indicated that current assessment techniques for wood chemistry, using wood cores extracted at breast height, are effective.

Natural variation in the wood chemistry of *E. globulus* was investigated using NIR predictions. Physical wood traits, growth and decay were also measured. Genetic variation was found for lignin content and decay, with significant locality differences. The only trait to have significant family within locality variation was acid-soluble lignin content, giving a high narrow-sense heritability (0.51 ± 0.26). Family means heritabilities were high for lignin content, extractives content and decay (0.42 - 0.64). This suggested that traditional breeding may be used to improve these traits. Furthermore, the chemical wood traits were highly correlated with each other and with density and microfibril angle. This indicated that selection for the breeding objective traits (basic density, pulp yield and volume) could concurrently produce favourable states in wood chemistry, decay resistance and fibre properties.

Having identified natural genetic variation in lignin, molecular variation in the lignin biosynthesis gene, cinnamoyl CoA reductase (*CCR*), was explored, firstly for its effect on wood properties; and secondly for its use in phylogenetic reconstruction at the sectional level of *Eucalyptus*. Segregation and quantitative trait loci (QTL) approaches were used to identify the effect of an amino acid substitution at a highly conserved position in CCR, on lignin content, lignin composition (syringyl/guaiacyl ratio) and wood density. The amino acid substitution had no significant influence on these traits nor did *CCR* collocate with any of the QTL for growth and density found in that cross. The high levels of sequence variation found for *CCR* in *E. globulus* suggested it may be useful for testing the monophyly of *Eucalyptus* sections *Exsertaria* and *Latoangulatae*, using section *Maidenaria* as an outgroup. *Latoangulatae* and *Maidenaria* were polyphyletic or paraphyletic, while *Exsertaria* species formed a clade but included a single *Latoangulatae* species. Analysis of intragenic recombination, a confounding factor when using nuclear genes for phylogenetic analysis, identified two events involving species from different sections. The occurrence of intragenic recombination may explain the anomalous positions of some species within the phylogenetic tree, and also suggests that levels of linkage disequilibrium will be low, which has implications for association studies. The high level of *CCR* sequence variation between and within *Eucalyptus* species suggests that molecular variants may be found in natural populations that will allow selection for improved lignin profiles.

PREFACE

Eucalyptus globulus is a temperate hardwood species native to south-east Australia. Its wood quality is ideal for pulp production, and because of this it is now grown in temperate areas throughout the world. During kraft pulping, lignin and extractives are chemically removed from the cellulose fibres (Smook, 1992). If high quality paper is the desired end product, the pulp must then be bleached to remove the residual lignin (Smook, 1992). These processes have been implicated in contributing to environmental damage. Pulping trees that have less lignin and extractives, or have lignin with a composition that is more easily removed (greater ratio of syringyl to guaiacyl subunits), will be beneficial to the process from both an industrial and environmental viewpoint.

The genetic parameters for lignin and extractives contents in *E. globulus* have been studied very little. Genetic variation is required for the improvement of these traits in breeding programs. The identification of the relationships among lignin and extractives and the other wood traits will also indicate how multiple traits are affected during the selection of superior trees. To quantify these parameters, large numbers of trees from the breeding population need to be screened, however, for chemical wood traits this has been impeded by the enormous cost and time-consuming nature of laboratory measurements. As such, indirect methods of phenotyping have been explored. Near infrared reflectance (NIR) analysis offers a low cost alternative that has been found to reliably predict other chemical wood traits in *Eucalyptus* (Wright et al., 1990; Schimleck et al., 1999; Raymond and Schimleck, 2002). This technique is likely to also provide a reliable method for predicting lignin and extractives contents in *E. globulus*, and will assist efforts in phenotyping the breeding population to assess genetic variation in these traits, and also correlations between them.

Once the genetic variation in the chemical wood properties is evaluated, the cause of this variation can be examined at the DNA level. In particular, lignin is a good candidate for investigation as its biosynthetic pathway is better understood (Baucher et al., 1998) than the pathways of other chemical wood traits. Lignin biosynthesis

genes and transcription factor genes have been cloned in *Eucalyptus*, genetic linkage maps constructed, quantitative trait loci found for lignin variation and key genes down-regulated in transgenic plants to assess gene function (Lacombe et al., 1997; Piquemal et al., 1998; Gion et al., 2000; Chabannes et al., 2001; Gion et al., 2001; O'Connell et al., 2002; Baucher et al., 2003; Goujon et al., 2003). Cinnamoyl CoA reductase (*CCR*) has been identified as a gene likely to influence lignin variation. The *CCR* gene sequence has a high level of variation in *E. globulus*, some of which affects the amino acid encoded and therefore potentially affects protein function (Poke et al., 2003; Appendix 1). Linking molecular variation with phenotype will not only increase our understanding of lignin development but will also provide a molecular marker which can be used as a selection tool in tree breeding. The identification of a high level of sequence variation in a single copy nuclear gene such as *CCR* also suggests it may be a useful tool that can be harnessed for answering phylogenetic questions in *Eucalyptus*. As the exploration of evolutionary relationships using chloroplast and nuclear encoded ribosomal DNA reach the limit of the information they can provide in this area, low copy number nuclear genes will become more important and the discovery of suitable candidates and their analysis will be valuable for shedding light on this complex area.

This study begins by summarising the genomic research that has been conducted in *Eucalyptus* to date, including the molecular understanding of wood quality and its place in worldwide research (Chapter 1). The following chapters detail the development of techniques that can be used for screening trees for lignin, cellulose and extractives contents, based on NIR analysis (Chapters 2 and 3). Within-tree variation in wood chemical composition is subsequently explored in Chapter 4, to determine if current sampling strategies employed in this study are effective. Using the information and applications provided by previous chapters, the genetic variation in chemical wood traits and their relationship to physical wood traits and decay are evaluated using several localities covering the range of *E. globulus* (Chapter 5). The following chapters explore molecular variation in the lignin biosynthesis gene *CCR*, firstly to determine whether an amino acid substitution at a highly conserved sequence position contributes to lignin variation (Chapter 6), and secondly to investigate the phylogenetic utility of *CCR* for resolving sectional level relationships in *Eucalyptus* (Chapter 7).

CHAPTER 1

Genomic research in *Eucalyptus*

This chapter has been published as: Poke F.S., Vaillancourt R.E., Potts B.M., Reid J.B. (2005) Genomic research in *Eucalyptus*. Genetica 125: 79-101.

Introduction

In an era of rapidly advancing, high-throughput molecular techniques and bioinformatics, large scale genomic studies are now performed routinely in many laboratories around the world. Extensive research has been conducted on mammalian genomes, and in plants, model species such as *Arabidopsis thaliana* and those of economic importance including *Oryza sativa* (agriculture) and *Populus trichocarpa* (forestry) are becoming increasingly well characterised (Brunner et al., 2004). The genomes of these species have been sequenced and will provide a valuable tool for gene discovery, determining gene function and interaction, and for understanding ecological responses, evolutionary processes and gene distribution across populations (Brunner et al., 2004). Angiosperms are thought to share many of the same genes and biochemical pathways, and therefore, some of the findings made in these model systems may be transferable to less well characterised species such as those of the genus *Eucalyptus*. *Eucalyptus* is a key forestry genus grown world-wide for solid timber as well as pulp and paper production (Doughty, 2000; Potts, 2004). However, despite its ecological importance to Australia and economic worth to world forestry, genomic research in *Eucalyptus* lags behind its cropping counterparts and forestry genera of the northern hemisphere such as *Populus*. This review collates the genomic research that has been conducted in *Eucalyptus*.

The genus *Eucalyptus* is a member of the *Myrtaceae* family and is composed of more than 700 species (Brooker, 2000). The latest taxonomic classification of eucalypts has included *Corymbia* and *Angophora* as subgenera of the *Eucalyptus* genus (Brooker, 2000), although some taxonomists recognise these as genera.

Recent DNA studies have supported their status as separate genera, showing *Angophora* and *Corymbia* to be more genetically similar to each other than to the other eucalypts (Udovicic et al., 1995; Udovicic and Ladiges, 2000; Steane et al., 2002). For the purpose of this review, we treat eucalypts in the broad sense and have included *Corymbia* and *Angophora* along with *Eucalyptus*.

Eucalypts are native to Australia and the offshore islands to its north. They occur over a wide range of environmental conditions, from sea level to alpine treeline and high rainfall to semi-arid areas, and vary in form, from shrubs to giant trees (Williams and Woinarski, 1997). Eucalypts are renown for their fast growth rate, straight form, growth ability in a wide variety of climates and soils, and their good wood quality for both solid wood products and pulp production (Eldridge et al., 1993; Doughty, 2000). Considerable inter- and intra-specific genetic variation occurs in these traits and their wide adaptability has made *Eucalyptus* an ideal plantation tree world-wide (Eldridge et al., 1993).

Extensive plantations of *Eucalyptus* are grown not only across Australia, but also throughout Asia, South America, Southern Europe and Africa, with at least 12 million hectares planted (Turnbull, 1999). *Eucalyptus grandis*, *E. globulus* and *E. camaldulensis* and their hybrids are the most important plantation species, comprising about 80% of the plantation area (Potts, 2004). *Eucalyptus nitens*, *E. saligna*, *E. urophylla*, *E. tereticornis*, *E. deglupta*, *E. pilularis* and *Corymbia citriodora* are also commonly grown (Potts, 2004). These are all classified within subgenus *Symphyomyrtus* except species of *Corymbia*. For pulp production and increasingly for solid wood *E. grandis*, *E. urophylla* and their hybrids are the most favoured in tropical and subtropical regions with *E. globulus* favoured in temperate regions (Potts, 2004). Traditional breeding programs have aimed to exploit the natural genetic variation in eucalypts, and their ability to hybridise, to identify genotypes which produce more favourable phenotypes. These are directed at economically and environmentally important characters such as growth, wood properties, pest resistance and stress tolerance. Hybrids have often been found to have a more favourable combination of traits than the parent species, serving to increase the already enormous genetic diversity in eucalypts (Potts and Dungey, 2004). Within eucalypt subgenera, hybridisation has been shown to be a relatively

common phenomenon between species, although, it does not readily occur between species of the different subgenera (Griffin et al., 1988).

Genomic research covers all aspects of the structure, function and sequence composition of the genome. Characterisation of the size and structure of the genome, including both nuclear and organellar genomes, has involved estimates of DNA content and the generation of linkage maps with the positioning of markers and quantitative trait loci (QTL). The isolation of genes and determination of gene function has been another major aspect of studies and has involved the cloning and identification of genes, gene expression profiling, collocation of genes with quantitative trait loci, association of DNA polymorphisms with phenotype and the manipulation of genes through transgenic plants. Studies into genome sequence composition have aimed to identify regions of both conservation and variation between individuals and species. This review will detail research that has been conducted in these areas in *Eucalyptus*.

Genome size and structure

Eucalypts are diploid plants with a haploid chromosome number of 11 (Eldridge et al., 1993; Potts and Wiltshire, 1997). The size of the genome has been estimated for several eucalypt species and their hybrids by Grattapaglia and Bradshaw (1994), using flow cytometry and comparison to chicken erythrocytes. They estimated a haploid genome size ranging from 370-700 million base pairs (Mbp). *Symphyomyrtus* species had on average a haploid genome size of 650 Mbp and species within the same section were found to have similar DNA contents, with *E. globulus* and *E. dunnii* at the lower end of the scale (530 Mbp) and *E. saligna* at the higher end (710 Mbp) (Grattapaglia and Bradshaw, 1994). *Corymbia* species were reported to have a haploid genome size of around 380 Mbp, which is much smaller than the other eucalypts (Grattapaglia and Bradshaw, 1994). No *Angophora* species have been studied. Hybrids were found to have an intermediate DNA content between the two parent species with no evidence of polyploidy detected (Grattapaglia and Bradshaw, 1994). Pinto et al. (2004) recently re-estimated the DNA content of *E. globulus* using a plant internal standard (*Lycopersicon esculentum* nuclei) and reported an average of 644 Mbp, which is larger than that

estimated by Grattapaglia and Bradshaw (1994). They suggested this difference could be due to different laboratory methods, the internal standard used or the tissue used, and it must be considered that it is not until the genome is completely sequenced that the actual size can be determined.

Regardless of variation in estimates, it is evident that the genome size of eucalypts is a little larger than some of the plant species which have recently had their genomes sequenced. *Arabidopsis thaliana* has one of the smallest plant genome sizes of 125 Mbp (The Arabidopsis Genome Initiative, 2000) and the two varieties of rice sequenced *Oryza sativa* L ssp. *indica* (466 Mbp) (Yu et al., 2002) and *Oryza sativa* L ssp. *japonica* (420 Mbp) (Goff et al., 2002) have comparable genome sizes to the recently sequenced *Populus trichocarpa* (approximately 473 Mbp) (http://genome.jgi-psf.org/Poptr1/Poptr1.info.html). In comparison to gymnosperms such as *Pinus,* eucalypts have a considerably smaller estimated genome size. Bogunic et al. (2003) estimated the genome size of five *Pinus* species using flow cytometry and identified interspecific variation with genome sizes ranging from 20,830-26,920 Mbp. This enormous size difference is attributed to large regions of repetitive DNA in pines (Ahuja, 2001).

Organellar genome structure and inheritance has also been investigated in eucalypts. Studies have found the chloroplast to be maternally inherited in eucalypts (Byrne et al., 1993; McKinnon et al., 2001b). Recently the chloroplast genome was sequenced for *E. globulus* with a size of 160,286 base pairs found (Steane, 2005). By comparing the *E. globulus* chloroplast genome sequence with that of other sequenced species including *Nicotiana tabacum* and *Oenothera elata*, very high homology was identified in the coding regions between these species and high divergence in the intragenic regions, which were also a source of microsatellites (Steane, 2005). Eucalypt mitochondria have been less well characterised. Vaillancourt et al. (2004) found the mitochondria to be maternally inherited in *E. globulus*, however, the size of the eucalypt mitochondrial genome is still unknown.

The nuclear genome of eucalypts has been represented by maps generated through linkage analysis of DNA based polymorphic markers (Shepherd and Jones, 2005). These markers include microsatellites or simple sequence repeats (SSRs), random

amplified polymorphic DNA (RAPD), amplified fragment length polymorphisms (AFLPs) and restriction fragment length polymorphisms (RFLPs) (Thamarus et al., 2002). Some of these markers are dominant, such as RAPDs and AFLPs. Microsatellites and RFLPs on the other hand are codominant markers and are generally more informative as they are multiallelic. Although dominant markers are relatively cheap they have limited transferability, whereas codominant markers have a much broader transferability and are potentially more informative in crosses with up to four alleles segregating such as outcross F_2 and pseudo-testcrosses (Bundock et al., 2000).

DNA markers that can be amplified across individuals and species are particularly useful as they can facilitate the construction of linkage maps in multiple species and allow for comparison between these maps. The transferability of markers across *Eucalyptus* is likely to be determined by how closely related the individuals or species are. Analysis of RAPD markers across individuals from the same population of *E. urophylla* showed a moderate 64% to be transferable and would therefore be informative among individual tree linkage maps (Brondani and Grattapaglia, 1996b). The transferability of microsatellite markers across species of *Symphyomyrtus* has been studied extensively (Byrne et al., 1996; Van Der Nest et al., 2000; Glaubitz et al., 2001; Brondani et al., 2002; Marques et al., 2002) with high rates identified. Ninety percent of microsatellites could be amplified in both *E. grandis* and *E. urophylla* (Brondani et al., 2002) and 78% of markers developed in *E. grandis* and *E. urophylla* could be transferred to *E. globulus* and *E. tereticornis* (Marques et al., 2002). Byrne et al. (1996) examined microsatellite transferability across subgenera and identified that 50% of microsatellite loci developed in *E. nitens* could be amplified in species from subgenus *Monocalyptus* (*E. sieberi, E. marginata*). The transferability of markers across species suggests that a consensus map is possible for *Eucalyptus* (Brondani et al., 2002). Studies have also shown a relatively high transferability of markers between *Eucalyptus*, *Corymbia* and *Angophora* (Jones et al., 2001; Steane et al., 2001). Fourteen microsatellite loci isolated from *C. variegata* had complete cross-species amplification in *Corymbia* and 71% amplified in at least one *Eucalyptus* species (Jones et al., 2001). In summary, these studies showed primer sequences for marker amplification to be more highly conserved between species of the same subgenera than across

subgenera, but also that some conservation exists between the more distantly related *Corymbia* and *Angophora* to *Eucalyptus*. Cross-species PCR amplification does not indicate that microsatellite markers will be polymorphic across the genera and it is likely that many more microsatellites will need to be developed in order to study linkage synteny between the eucalypts (see below).

In eucalypts linkage maps have been constructed using both inter- and intra-specific crosses. These linkage maps are summarized in Table 1 according to the cross type used to generate these maps. One of the first maps was generated by Grattapaglia and Sederoff (1994) using a pseudo-testcross mapping strategy for a *E. urophylla* x *E. grandis* interspecific cross and RAPD markers. This approach involves crossing two individuals for which many polymorphic markers will be heterozygous in one parent and homozygous in the other, and will therefore segregate in a 1:1 ratio in the F_1 progeny much like a testcross (Grattapaglia and Sederoff, 1994). The pseudo-testcross strategy was identified as an efficient means for generating individual linkage maps for highly heterozygous organisms and is now commonly used. However, with this strategy the markers are not shared between the parents and therefore maps can not be compared (Grattapaglia and Sederoff, 1994) unless markers segregating from both parents are also mapped. While interspecific crosses have been widely used for generating linkage maps in eucalypts (Grattapaglia and Sederoff, 1994; Verhaegen and Plomion, 1996; Marques et al., 1998; Gion et al., 2000; Myburg et al., 2003), for species such as *E. globulus* which are mainly grown as a pure species, intraspecific crosses are generally required. Two linkage maps have been published using intraspecific F1 pseudo-testcrosses including both inter- (Thamarus et al., 2002) and intra-provenance (Bundock et al., 2000) crosses of *E. globulus*.

The transferability of molecular markers between species has made possible the study of linkage synteny between species. Linkage synteny is a measure of the similarity between genomes and is measured as the conservation of both position and order of markers and genes. It is important since synteny can be used to transfer knowledge gained from linkage maps for one species or cross to another. Within a single eucalypt species Brondani and Grattapaglia (1996a) showed RAPD markers to have a high conservation of linkage and locus order among individual trees from

the same population of *E. urophylla*. Linkage synteny has also been examined across species with colinear syntenic linkage groups established for *E. urophylla* and *E. grandis* (Brondani et al., 2002) and for four *Symphyomyrtus* species (*E. grandis*, *E. urophylla*, *E. tereticornis*, *E. globulus*) (Marques et al., 2002). These results further suggested that a consensus map, formed from the integration of information from individual maps, may be possible for the *Eucalyptus* species. A consensus map could be used to increase marker density in certain regions of species specific maps and to map species that have low levels of polymorphism in the parental lines (Marques et al., 2002)

Segregation distortion of markers is another important concept to be considered when studying the structure of the eucalypt genome. The transmission of parental alleles in advanced hybrid generations can be followed using comparative genome maps (Myburg et al., 2004). This approach can be used to explore the mechanism behind the genetic incompatibility that is seen between some eucalypt species (Potts and Dungey, 2004). Myburg et al. (2004) conducted a whole genome analysis of barriers to introgression between two species from different taxonomic sections, *E. grandis* and *E. globulus,* isolated by prezygotic and postzygotic barriers. 1354 AFLP markers were evaluated in linkage maps generated for the backcross parents and an F_1 individual, and these showed much colinearity and suggested that the postzygotic isolation was not due to large chromosomal rearrangement (Myburg et al., 2004). 27.7% of marker loci exhibited transmission ratio distortion and these were located in distinct regions of the parental maps with their transmission biased toward either of the two parent species (Myburg et al., 2004). This finding suggested strong genic incompatibilities in the backcross pedigree. Other phenomena important to genome structure such as small translocations of sequence from one chromosome to another or inversions of DNA sequence have yet to be studied in eucalypts.

Linkage maps are not only a representation of the genome but also provide a means of studying the genetic control of complex traits. Regions of the genome which affect the variation in a quantitative trait can be identified through QTL analysis. Using either the framework markers on a linkage map or their intervals, variation in phenotypic measurements can be linked to positions on linkage maps, essentially

identifying regions of the chromosome influencing that trait. Although this technique is a very powerful tool it does not pinpoint the gene responsible for the quantitative trait. The confidence interval of a QTL is usually around 10-20 cM, which is in the order of 4.5-9 Mb in *Eucalyptus* and may be the equivalent of hundreds of genes (Gion et al., 2000). If a gene has been mapped to this region and is thought to be a candidate for altering that particular phenotype, either through its involvement in biochemical or regulatory pathways, this is the first level of evidence for that gene being responsible for influencing variation in that trait (Thamarus et al., 2002; See section on gene isolation and function). QTL studies have been conducted in a large range of eucalypt species (Table 1), particularly *E. grandis*, *E. globulus* and *E. urophylla* as the leading commercially grown species or their hybrids, and have predominantly focussed on commercially important traits such as wood properties, growth, propagation, and disease and insect resistance.

To fully exploit the information gained from QTL studies and to extrapolate it across populations and species, it is important to know if QTL are stable across ages and environments. This information is limited in eucalypts and two major studies have shown alternative results for the QTL being examined. Campinhos et al. (1996) found QTL detected for volume growth in a *E. urophylla* x *E. grandis* interspecific cross to be unstable in variable genetic backgrounds and to also be compounded by an age effect. The usefulness of these particular QTL will therefore be reduced outside this cross and the dependency of the QTL on age suggests the identification of the underlying genetic control of phenotype will be more difficult. A study conducted by Verhaegen et al. (1997) using a *E. urophylla* x *E. grandis* F_1 hybrid population found 70% of wood density, stem growth and stem form QTL to be stable across ages, with only a few detected at a single age. One of the reasons for instability may be the imprecision in estimating QTL effect due to small progeny size (Beavis, 1998). The first report of QTL stability across species was made by Marques et al. (2002). QTL influencing vegetative propagation traits were found to be located on homeologous linkage groups in *E. grandis*, *E. urophylla*, *E. tereticornis* and *E. globulus*. Together these results suggest that QTL stability will be heavily reliant on age and genetic background, and transfer of QTL between species will likely be dependent on the nature of the QTL in question and the species examined.

Table 1. Genetic linkage maps and QTL analyses in *Eucalyptus*.

Cross type	Markers used	QTL analysis	Reference
E. camaldulensis x *E. camaldulensis*	RAPD, RFLP, SSR		Agrama et al. (2002)
E. camaldulensis x *E. globulus*	RAPD		Li et al. (1999)
E. camaldulensis x *E. grandis*	RAPD	Salt tolerance	Dale et al. (2000)
E. erythronema x *E. stricklandii*	RAPD, SSR, ISSR	Propagation traits	M. Sedgley, personal communication
E. globulus x *E. globulus*	RFLP, RAPD, SSR, AFLP, isozyme, EST, candidate gene	Tree height, wood density, growth, *Mycosphaerella* resistance, wood and fibre traits	Song and Cullis (1992), Mitchelson et al. (1997), Bundock et al. (2000), Thamarus et al. (2002), Bundock (2003), Freeman et al. (2003), Thamarus et al. (2004)
E. globulus x *E. tereticornis*	AFLP	Vegetative propagation traits	Marques et al. (1998, 1999)
E. grandis x *E. globulus*	AFLP, candidate gene	Wood properties	Myburg et al. (2003), Kirst et al. (2004)
E. grandis x *E. grandis*	RAPD	Rust resistance, wood splitting, phenotypic anomaly, early flowering locus	Barros et al. (2002), Missiaggia et al. (2002), Junghans et al. (2003), Nogueira et al. (2004)
E. grandis x *E. urophylla*	RAPD, SSR, candidate gene	Christmas beetle resistance, foliar oil composition, biomass traits (height, diameter, wood density, bark thickness), vegetative propagation, growth, stem growth, stem form, lignin content and lignin composition	Bradshaw and Grattapaglia (1994), Grattapaglia and Sederoff (1994), Grattapaglia et al. (1995), Shepherd et al. (1995), Campinhos et al. (1996), Grattapaglia et al. (1996), Verhaegen and Plomion (1996), Chaparro et al. (1997), Verhaegen et al. (1997), Brondani et al., (1998), Mitchelson et al. (1999), Shepherd et al. (1999), Gion et al. (2001), Brondani et al. (2002)
E. gunnii x *E. globulus*	RAPD	Frost tolerance	Vaillancourt et al. (1994)
E. nitens x *E. nitens*	RFLP, RAPD, isozyme	Frost tolerance, seedling height, leaf area, wood traits	Byrne et al. (1995, 1997a, 1997b), http://www.ffp.csiro.au/tigr/atrnews/atrn07/atrnews7_06.htm
E. urograndis x *E. globulus*	RFLP, SSR	Frost tolerance	Fullard and Moran (2003)
E. urophylla x *E. tereticornis*	RAPD, SSR, RFLP, EST, candidate gene		Gan et al. (2003)
E. urophylla x *E. urophylla*	RAPD, SSR	Bacterial wilt resistance	S. Huang, personal communication

Large mapping and QTL projects have been initiated to create more informative, high density maps for more species, using larger progeny sizes and looking at a wider range of commercially important traits for QTL identification. A collaborative project has been initiated between European partners UMR CNRS-UPS (France), CIRAD (France), ENCE (Spain), and RAIZ (Portugal) aimed at identifying and mapping candidate genes and QTL responsible for variation in wood properties (Paux et al., 2003). SSR maps are being created and gene-markers developed and mapped (Paux et al., 2003). The Genolyptus project, a consortium of seven Brazilian universities, twelve companies, Embrapa and RAIZ is using 24

connected full-sib families involving several eucalypt species and hybrids for mapping (Grattapaglia, 2002). QTL analysis will be undertaken for wood properties, growth, flowering and disease resistance (Grattapaglia, 2002). These large scale studies will help address the issues of the synteny of DNA markers across individuals and species, and the stability of detected QTL in various environments, ages, crosses and species, aiming to generate consensus maps.

Gene isolation and function

Currently the number of genes sequenced in eucalypts is limited, particularly in comparison to other species such as *Arabidopsis thaliana*. To date, around 4000 *Eucalyptus* sequences (excluding ribosomal and chloroplast sequences) can be found in the public database NCBI GenBank. The genes isolated and characterised in *Eucalyptus* are mainly those involved in a few important developmental processes within the plant. A large focus has been placed on genes involved in wood formation, due to its economic importance, and the majority of this work has been on genes involved in lignin biosynthesis (Piquemal et al., 1998). Lignin is the major component of wood after cellulose and has a negative impact on chemical pulping. The lignin biosynthesis pathway has been well characterized in other species (Grima-Pettenati and Goffner, 1999; Baucher et al., 2003) and many of the genes for the pathway enzymes, and also for transcription factors which regulate the expression of these genes (e.g. *MYB* genes and *LIM* homeobox genes), have now been cloned and expression patterns analysed in *Eucalyptus* (Table 2). Cloning of the majority of genes in a pathway opens the way for more detailed expression studies which can elucidate movements of substrates through pathways, identify which genes are imperative to the process and increase our understanding of why certain responses are observed when genes are up-or down-regulated.

Table 2. Lignin biosynthesis and regulatory genes cloned in *Eucalyptus*.

Gene	Species	NCBI GenBank Accession Number	Reference
aldOMT	*E. camaldulensis*		Ho et al. (2002)
CAD	*E. gunnii*	X75480, X65631, X88797	Boudet et al. (1993), Feuillet et al. (1993), Grima-Pettenati et al. (1993), McKie et al. (1993), Hawkins et al. (1994), Hawkins and Boudet (1994), Goffner et al. (1995, 1998), Lauvergeat et al. (2002)
	E. globulus	AF109157, AF038561	De Melis et al. (1999)
	E. botryoides	D16624	Hibino et al. (1994)
	E. saligna	AF294793	Endt and Pasquali, unpublished[a]
CCoAOMT1,	*E. gunnii*	Y12228, X74814	Poeydomenge et al. (1994)
CCoAOMT2	*E. globulus*	AF168778-AF168780	De Melis et al., unpublished[a]
CCR	*E. gunnii*	AJ132750, X79566, X97433	Goffner et al. (1995), Lacombe et al. (1997, 2000)
	E. globulus	AH011638	Poke et al. (2003)
	E. saligna	AF297877	Zago and Pasquali, unpublished[a]
COMT1, COMT2	*E. gunnii*		Hawkins et al. (2003)
	E. globulus	AF168776, AF168777, AF046122	De Melis et al., unpublished[a]
F5H	*E. gunnii*	AJ249093	Rech et al., unpublished[a]
laccase	*E. globulus*	AW191310	Bossinger and Leitch (2000)
LIM homeobox	*E. globulus*		Bossinger and Leitch (1999)
myb1, myb2	*E. gunnii*	AJ576023, AJ576024	Goicoechea et al., unpublished[a]
PAL	*E. globulus*	AF167487	De Melis et al., unpublished[a]
4CL	*E. gunnii*	AJ244010	Rech et al., unpublished[a]
	E. globulus	E64536, E64537, AW191302	Bossinger and Leitch (2000), Kajita and Kawabata, unpublished[a]

[a] GenBank citation

Flowering genes have been extensively studied in plants such as *Arabidopsis* and *Antirrhinum* (Zik and Irish, 2003) but only to a small extent in *Eucalyptus*. To assist in the understanding of floral morphogenesis and to determine how eucalypt flower structure and development compares to other species, the homologues of many of the flowering genes have been isolated and characterized predominantly in *E. globulus* and *E. grandis*. The flowering genes that have been cloned in *Eucalyptus* are detailed in Table 3. *Eucalyptus* homologues have been found for *Arabidopsis LEAFY* and *AP1*, and *Antirrhinum FLORICAULA* which are involved in the first step of flower development, creation of the floral meristem (Kyozuka et al., 1997; Southerton et al., 1998b). The two eucalypt homologues of *Arabidopsis AP1* (*EAP1* and *EAP2*) were found to have 60-65% homology to their *Arabidopsis* counterpart (Kyozuka et al., 1997). The high homology and similar expression patterns of the eucalypt genes to the *Arabidopsis* flowering genes suggested comparable controls of flower development between the two species despite the modified floral structure of eucalypts (Kyozuka et al., 1997; Southerton et al., 1998b).

Table 3. Flowering genes cloned in *Eucalyptus*.

Gene	Species	NCBI GenBank Accession Number	Reference
Floral meristem identity genes *EAP1, EAP2*	*E. globulus*	AF305076, AF305696, AF306349	Kyozuka et al. (1997)
Floral meristem identity genes *ELF1, ELF2*	*E. globulus*	AF034806	Southerton et al. (1993, 1998b)
MADS-box genes *egm1, egm2, egm3*	*E. grandis*	AF029975-AF029977	Southerton et al. (1998a)
SVP-like floral repressor mRNA	*E. occidentalis*	AY273873	Watson and Brill, unpublished[a]
SVP-like floral repressor mRNA	*E. grandis*	AY263809	Brill and Watson (2004)
TM3-like MADS-box gene	*E. globulus*	AF086642	Decroocq et al. (1999)
Two *SOC1*-like floral activators mRNA	*E. occidentalis*	AY273872, AY273874	Watson and Brill, unpublished[a]
Two *SOC1*-like floral activators mRNA	*E. grandis*	AY263807, AY263808	Watson and Brill (2004)

[a] GenBank citation

Many of the genes that have been found to direct floral development contain a highly conserved MADS-box domain, which is involved in DNA binding of a transcription factor (Southerton et al., 1998a). Several of these MADS-box genes have been isolated in *Eucalyptus* (Southerton et al., 1998a; Decroocq et al., 1999). Southerton et al. (1998a) identified three single copy MADS-box genes, two of which had strong homology with other plant MADS-box genes that mediate between the floral meristem and organ-identity genes (85% and 63%) and the third most homologous to genes which regulate organogenesis of the second and third floral whorls (62%). This is a good example of how gene homology can be used to infer the function of genes. Studies into flowering have implications for genetic engineering as the creation of sterile trees may be essential for avoiding the release of transgenes into native populations. On the other hand, the identification of genes which can be used to induce flowering will be important for the creation of early flowering trees which will assist in these research efforts (http://users.ox.ac.uk/~dops0022/project4.html).

The characterisation of genes involved in nutrition and stress tolerance will facilitate understanding of these important characters and help us understand why certain species are better suited to particular environmental situations. This area of eucalypt gene cloning is more recent and as a result these genes have been less well characterised. These genes mainly include nutrient transporter genes (Table 4). Salt

tolerance and nutrient uptake genes have recently been examined in different tissues of eucalypt clones that have varying salt tolerance (Faridah, 2004).

Table 4. Cloned genes involved in plant nutrition and stress responses in *Eucalyptus*.

Gene	Species	NCBI GenBank Accession Number	Reference
Calcineurin-like protein gene *EcCBL1*	*E. camaldulensis*	AF197334	Fairbairn et al., unpublished[a]
Calcineurin-like protein gene *EgCBL1*	*E. grandis*	AF197330	Fairbairn et al., unpublished[a]
Dehydrin gene	unknown		Ishige et al. (2004)
Fertilisation independent endosperm development protein mRNA	*E. gunnii*	AY150283	Danilevskaya et al. (2003)
Inward rectifying K+ channel gene *EcKT1;* mRNA *EcKT1-1*, *EcKT1-2*	*E. camaldulensis*	AF197331-AF197333	Fairbairn et al., unpublished[a]
Mitochondrial malate-dehydrogenase cDNA	*E. gunnii*		Poeydomenge et al. (1995)
Phosphate transporter gene	*E. camaldulensis* x *E. globulus*		Faridah (2004)
Potassium transporter gene	*E. camaldulensis* x *E. globulus*		Faridah (2004)
Sodium/hydrogen antiporter gene	*E. camaldulensis* x *E. globulus*		Faridah (2004)
Sulfate transporter gene	*E. camaldulensis* x *E. globulus*		Faridah (2004)
Tocopherol cyclase mRNA *sxd1*	*E. gunnii*	AY336944	Marque et al. (2003)
Two sodium-potassium symporters mRNA *HKT1*, *HKT2*	*E. camaldulensis*	AF176035, AF176036	Fairbairn et al. (2000)
Zinc transporter mRNA *EgZnT1*	*E. grandis*	AF197329	Fairbairn et al., unpublished[a]

[a] GenBank citation

Eucalypts are susceptible to a range of diseases (Keane et al., 2000). Identifying the genetic basis of disease resistance is an important component of research particularly aimed at plantation forestry. Breeding for resistant genotypes or using pesticides are currently the only methods for combatting eucalypt diseases (Chimwamurombe et al., 2001; Milgate et al., in press). Many genes play a role in a plant's defense system, however, currently only a single gene has been examined in *Eucalyptus*. The gene coding for the polygalacturonase-inhibiting protein (*PGIP*), which is produced as part of the plant's defense against disease, has been partially cloned in *E. urophylla*, *E. camaldulensis*, *E. nitens* and *E. saligna* (GenBank Accession No.s AF159168-159171; Chimwamurombe et al., 2001) and completely cloned in *E. grandis* (GenBank Accession No. AY445043; Bhoora et al., 2003). The *pgip* gene was highly conserved between the eucalypt species with 98-100% identity, although, the identity of the amino acid sequences with other plant *pgip*

genes showed varying levels of conservation (44-94%). It is proposed that disease tolerance may be improved through increasing the expression of this gene (Chimwamurombe et al., 2001). A locus for rust resistance has been found in *E. grandis*, although, the gene has not yet been identified and is currently the focus of cloning efforts (Junghans et al., 2003).

Expressed sequence tags (ESTs) are used extensively for gene isolation and identification, and for determining which genes are expressed in a particular tissue, at a particular time and environment (Hatey et al., 1998). An EST is a sequenced fragment of one end of a clone from a cDNA library (Hatey et al., 1998) and collectively they represent the expressed fraction of the genome. cDNA libraries have been made for a whole range of eucalypt species and tissues (Table 5) with wood formation and stress responses the two key processes primarily examined. Some of these studies have created cDNA libraries with a specific plant process in mind, other larger projects have encompassed multiple species, tissues and environmental constraints. ESTs can represent previously identified genes or novel sequences and may be identified through comparison to existing sequences in a database (annotation) (Sreenivasulu et al., 2002). Comparisons between expressed genes of different tissues in the same plant and of the same tissues in different species may give insight into developmental processes and why some species have more favourable properties such as a higher wood density or pulp yield.

Some large research projects have been established specifically for EST sequencing and identification. Saltgrow Pty Ltd (Australia) has constructed several cDNA libraries for *E. camaldulensis*, *E. grandis* and *E. globulus* using different tissues and different stress conditions producing a database of 5,000 ESTs from which a potassium channel gene was patented (G. Dale, personal communication). E. I. DuPont de Nemours and Company (USA) has contributed 2150 sequences to GenBank (Accession Nos CD667988-CD670137) including 1019 ESTs from *E. grandis* (60 of these from root tissue and 959 from leaf tissue including apical leaves [302], adult leaves [346] and combined leaf, petioles and stems [311]) and 1131 ESTs from *E. tereticornis* (542 from flowers and 589 from carpels). Agrigenesis (New Zealand) has constructed 20 different cDNA libraries from *E. grandis* using several different tissues including the xylem, phloem, roots,

reproductive organs, photosynthetic tissue, developing embryos and seedlings grown under different conditions. 182, 982 ESTs were sequenced and 951 of these are in GenBank (Strabala, 2004). They have also studied *E. dunnii* and *E. obliqua*. The FORESTs project (Brazil) has several cDNA libraries for *E. grandis* including: whole wood; flower, bud and fruits; young leaves; adult leaves; *Thyrinteina* infected leaves; young roots; adult roots; seedlings cultivated in the dark; and stems from young plants (L.E.A. Camargo, personal communication). cDNA libraries have also been made for *E. globulus*, *E. saligna*, *E. urophylla* and *E. camaldulensis* seedlings cultivated in the dark (L.E.A. Camargo, personal communication). An EST database has been established (https://forests.esalq.usp.br) with 123,889 ESTs of which 33,000 are unigenes (L.E.A. Camargo, personal communication). The Genolyptus project (Brazil) has 135,093 ESTs from a total of 29 different libraries derived from 13 different kinds of tissue, species or source DNA including: *E. grandis* mature leaf (8,067), young leaf (576), seedlings (14,502), treated seedlings with over 10 different treatments (17,369), flower (2,788), *Puccinia* (rust) infected leaves (7,200); *E. globulus* xylem (18,051); *E. pellita* xylem (11, 340); *E. urophylla* xylem (10,309); combined 10 species phloem (14,700) and root (2,304) cDNA libraries plus *E. grandis* shotgun genomic (10,289) and *E. grandis* bacterial artificial chromosome (BAC) end sequencing (17,598) libraries (D. Grattapaglia, personal communication). However, these large EST databases are not yet publicly accessible.

For most of the major EST projects annotation is the most time consuming part and is still in the early stages. The FORESTs project is conducting extensive data mining focussing on genes related to wood quality and resistance to abiotic and biotic stresses including: gene and allele identification related to tolerance to environmental stress and insect, disease, herbicide and weed damage (Alves et al., 2005); pathogenesis-related proteins (Furtado et al., 2004); herbicide resistance (Velini et al., 2004); phloem retranslocation of boron and enzymes related to sorbitol and mannitol synthesis and breakdown (Domingues et al., 2004); plant growth hormone genes (Mori et al., 2004); ABA biosynthesis genes and tissue expression (Guerrini et al., 2004); proline production genes (Trigueiro et al., 2004); genes for enzymes involved in terpenoid synthesis (Wilcken et al., 2004); and non hormonal growth genes (Zimback et al., 2004). Both the FORESTs and Genolyptus

projects have identified the "complete" set of 21 lignin genes. Agrigenesis are now using their EST database to direct a major effort towards promoter cloning (Strabala, 2004).

Table 5. cDNA library, EST and microarray resources for genomic research in *Eucalyptus*.

Species	Gene expression analysis	Reference
E. globulus	libraries for mature *vs.* juvenile wood, tension *vs.* opposite wood on microarrays, subtracted cambial genes (bark side, no phloem tissue minus mature leaves), 93 ESTs with 43 annotated; differential display and subtractive hybridisation used to identify root primordia formation genes; cDNA libraries for drought and cold stressed seedlings	Bossinger and Leitch (2000), Pelosi (2002), J. Hamill, personal communication, http://www.scsv.ups-tlse.fr/root/equipes/regulation/equipe_en.php; GenBank Accession No.s AW191301-AW191393; GenBank Accession No.s AJ697752-AJ697760, AJ697762-AJ697764
E. grandis	xylem cDNA library with 555 ESTs	GenBank Accession No.s CB967505-CB968059
E. gunnii	normalised subtractive hybridisation libraries for xylem *vs.* leaves, xylem *vs.* phloem, phloem *vs.* xylem; frost tolerance subtractive cDNA library and expression profiles of 3000 cDNA clones under cold, osmotic and salt stress	Keller et al. (2004), Paux et al. (2004), http://www.scsv.ups-tlse.fr/root/equipes/regulation/equipe_en.php; GenBank Accession No.s AJ627645-AJ627868
E. microcorys	differential display reverse transcriptase for genes down-regulated upon elevated salt levels	Chen and De Filippis (2001)
E. nitens	5000 xylem microarray examining vertical stems, branches; cDNA libraries for cambial scrapings (stem-side only), growing leaves, outer sapwood, induced tyloses; cDNA libraries for drought and cold stressed seedlings	G. Bossinger, personal communication, J. Hamill, personal communication, http://www.ffp.csiro.au/tigr/atrnews/atrn07/atrnews7_06.htm
E. urograndis x *E. globulus*	microarrays used to screen extreme cold tolerance phenotypes comparing cold hardened and unhardened	Fullard and Moran (2003)
Symphyomyrtus species	cDNA libraries with 50 gene tags provided for flowering regulation, salt stress	Sawbridge et al. (1999)
Multiple species	ESTs from differentiating xylem, leaves; cDNA microarrays comparing juvenile/mature wood, wild type/transformants	Sato et al. (2001)

Once a gene has been identified, its expression can then be analysed to determine if it is increased or reduced during a developmental process and under what conditions this occurs (Eisen et al., 1998). This allows the key genes in a developmental process to be identified and provides targets for genetic manipulation or molecular breeding. Microarray analysis has been a primary technique used to study gene expression (Eisen et al., 1998). A set of previously isolated genes, ESTs or oligonucleotides are spotted onto arrays to which labelled cDNA from a particular tissue can be hybridised (Duggan et al., 1999; Thomas and Klaper, 2004). From the

resulting label intensity of the spot, it can be determined if the gene in question has been up- or down-regulated, or if its expression has remained the same (Thomas and Klaper, 2004). It must be remembered, however, that this technique is limited to the set of genes being tested and is also influenced by which genes have already been isolated from the transcriptome (Draghici et al., 2003). Gene expression studies in eucalypts are summarised in Table 5.

Changes in gene expression during wood formation has been examined to provide insight into this complex process and to determine why different wood characteristics are produced in different individuals, genotypes and species. Using cDNA microarrays, Kirst et al. (2002) compared xylem gene expression between three different genotypes with contrasting wood and growth traits for both *E. globulus* and *E. grandis*. Only four genes were found to have significant differences in transcript level between species. These findings suggest that certain genes may be pivotal to the development of variable wood quality. A complementary technique to microarray analysis, cDNA-AFLP, which allows rapid, high-throughput screening for differentially expressed gene transcripts has been used to study gene expression during xylem formation in *E. grandis* x *E. nitens* hybrids (Ranik et al., 2004). Tissue samples were collected from different stages of wood development representing the variation in gene expression among different parts of the same tree and clonal replicates of the same tree (Ranik et al., 2004). A comparison of mRNA pools from the differentiating xylem, cambium, phloem and bark at each position will indicate which genes are essential to the development of wood and at what stages these are turned on or off (Ranik et al., 2004). Paux et al. (2004) examined xylem specific genes in *E. gunnii* using suppression subtractive hybridisation followed by transcript profiling using cDNA arrays. Eighty-one percent of the 224 independent ESTs obtained were preferentially expressed in the xylem with one-third of these for auxin signaling, cell wall biosynthesis and remodeling (Paux et al., 2004). Identification of the ESTs which did not match known function proteins shall be important for increasing understanding of xylem development. The Genolyptus project is also currently using microarrays for expression profiling of contrasting states and phenotypes (Grattapaglia, 2003).

Both gene and protein expression analysis have been used to identify the genetic control of tension wood formation. Tension wood forms in the zones of the tree held under tension and is characterized by a low lignin content, high cellulose content and low microfibril angle compared to normal wood (Plomion et al., 2001). It is a problem for solid timber products as it increases shrinkage during drying (Plomion et al., 2001). Jun-ichiro et al. (1997) have assessed the activity of 31 clones of expressed genes in the differentiating xylem during tension wood formation and the cDNA of genes which increase their expression during tension wood formation have been identified in *E. camaldulensis* (Aritome et al., 1999). A study conducted by Plomion et al. (2003) took another approach to discerning the mechanism of tension wood formation and analysed protein expression along the gradient from normal to tension wood in *E. gunnii*. Proteins were extracted from the differentiating xylem of a crooked tree and xylem samples classified as normal or tension wood through growth strain measurements. Twelve out of the 140 proteins studied showed significant association with growth strain, with seven less abundant in tension wood and five more abundant (Plomion et al., 2003). This suggests that these genes may play a role in the formation of tension wood.

A good example of the use of multiple genomic approaches for deciphering a complex process is in the study of the *E. globulus-Pisolithus tinctorius* ectomycorrhizae symbiotic relationship. Several genes thought to be involved in this process have been cloned and their expression analysed including alpha-tubulins (Diaz et al., 1996), a NADP-dependent isocitrate dehydrogenase (Boiffin et al., 1998) and a symbiosis regulated transcript called EgHypar (Nehls et al., 1998). Analysis of EgHypar suggested for the first time that gene expression of the eucalypt was altered by molecules from the ectomycorrhizal mycobiont (Nehls et al., 1998). EST sequencing and microarray analysis have been used to increase understanding of the morphological changes which occur in the roots during symbiosis. Tagu and Martin (1995) found 42% of the ESTs from their cDNA libraries to have identity to previously sequenced genes and suggested that these could act as markers for analysing gene expression. Voiblet et al. (2001) examined a set of 850 ESTs obtained from random clones or through suppression subtractive hybridization and identified 43% of these to be novel genes with the remainder found to be coding for proteins involved in gene/protein expression, cell wall

proteins, metabolic enzymes and signaling systems components (Voiblet et al., 2001). Free-living partners and symbiotic tissues have also been compared using cDNA arrays to identify symbiosis regulated genes with 17% of the genes analysed found to have altered expression (Voiblet et al., 2001). Duplessis et al. (2005) confirmed and extended these earlier observations by looking at gene expression during different stages of ectomycorrhiza development. Up-regulation of stress and defense-related genes was found to occur early in development, hormone metabolism genes mid-development, and protein synthesis and fate as well as nitrogen and carbon metabolism genes late in development. Genes coding for abscisic acid stress-related proteins were down-regulated during development. As with the study of Voiblet et al. (2001) no ectomycorrhiza specific gene was found. These studies reveal which of the already identified genes are important for the symbiotic interaction and the large number of novel ESTs found suggests there is much to learn about the genetic regulation of this process.

Collocation of a QTL with a gene or EST on a linkage map provides evidence that the gene may be responsible for part of the natural variation in a particular trait. However, it must be remembered that many genes can be found in the QTL region and this level of proof is affected by the number of genes that have been isolated and mapped, and by biochemical or microarray evidence of that gene being involved in the process. Gion et al. (2000) mapped lignification and symbiosis genes onto *E. grandis* and *E. urophylla* linkage maps using single strand conformation polymorphisms (SSCPs). Following QTL analysis for lignin content and composition, some of the QTL were found to collocate with genes involved in the lignin biosynthesis pathway and with transcription factors regulating that pathway, which suggests that these genes may have an effect on these traits (Gion, 2000; Gion et al., 2000). The *E. globulus* maps of Thamarus et al. (2002) contained 31 cambium-specific ESTs as well as 14 genes of known function involved in lignin and cell wall biosynthesis and floral development. QTL for pulp yield, cellulose content, fibre length and wood density were found to collocate with some of the mapped genes for cell wall biosynthesis, suggesting that these loci may contribute to variation in these quantitative traits (Thamarus et al., 2004). Not only is knowledge needed of the stability of a QTL across pedigrees and species, but also of the stability of QTL/candidate gene association. A project between UMR CNRS-

UPS, CIRAD, ENCE and RAIZ is currently exploring this question with candidate genes thought to affect wood properties being mapped in different *Eucalyptus* genetic backgrounds for association with QTL for wood density and pulp yield (Marques et al., 2003). Repeated association of a QTL with a candidate gene can further strengthen evidence of the role of that gene in causing variation in a quantitative trait.

A new method for identifying QTL/candidate gene associations involves the use of expression QTL, which Kirst et al. (2004) have used to study wood properties. Current QTL mapping studies do not have the fine resolution to identify individual genes underlying quantitative trait variation, as already discussed, however, by combining high-throughput genetic mapping with gene expression profiling this problem is being addressed (Kirst et al., 2004). AFLP markers were used to construct maps for the parents of a (*E. grandis* x *E. globulus*) x *E. grandis* cross. For 91 of the progeny variation in the transcript levels of 2,608 genes was determined through cDNA microarray analysis, and expression QTL were identified for 1,067 genes (Kirst et al., 2004). Association of expression QTL with QTL for wood properties allowed the identification of novel genes that underlie quantitative traits, such as the *RCI2* gene which explained one quarter of the variation in wood density. *RCI2* is thought to be involved in cell wall formation, however, it was not previously considered a candidate gene for wood density (Kirst, 2003).

To verify that a candidate gene is the one influencing a quantitative trait, association studies can be used to link the trait variation with single nucleotide polymorphisms (SNPs) within a gene (Thamarus et al., 2004). Association studies are relatively new in plants, especially in eucalypts. Often the SNP in question is not the one that affects the phenotype but it is in linkage disequilibrium (LD) with the causative SNP which may be down- or up-stream. LD is the non-random association of alleles at two or more loci on a chromosome. Although LD has been well studied in crop species such as maize, it has been examined little in eucalypts. However, it has been speculated that LD would be low due to the high outcrossing rates (Rafalski, 2002) in eucalypts. One of the first LD studies in eucalypts examined *CCR*, a gene specific to lignin biosynthesis, and found LD to decay fairly rapidly across the gene (G. Moran, personal communication).

Association studies are slowly being initiated in eucalypts as more genes are identified and isolated, and further evidence is obtained of the involvement of a candidate gene in influencing a quantitative trait. Provided with the knowledge of low LD in *CCR* and collocation of QTL for pulp yield and cellulose content with the *CCR* gene (Thamarus et al., 2004), the first association study in *Eucalyptus* has been conducted using *E. nitens* (G. Moran, personal communication). *CCR* has long been considered a candidate gene for variation in lignin content and composition (Piquemal et al., 1998; Gion et al., 2001), and it may also affect composite traits such as pulp yield (Thamarus et al., 2004). Twenty-five *CCR* SNPs, mostly in the introns or promoter, were assayed in 290 unrelated individuals of *E. nitens* (G. Moran, personal communication). Two of these SNPs were found to have a significant association with microfibril angle (the angle between the cellulose microfibrils in the cell wall and the axis) (G. Moran, personal communication).

The ultimate method for determining the function(s) of a gene is through the creation of transgenic plants where the gene of interest is down- or up-regulated. The increase or decrease in the transcript level of a gene may cause phenotypic changes that can be used to infer the role of that gene. Like other aspects of eucalypt genomics, the main traits under examination are wood formation, stress response and disease resistance (Table 6). A limitation to the use of this technique lies in the general recalcitrance of eucalypts to transformation and poor regeneration capacity. At this time, stable transformation has only been successful in a small number of species including *E. camaldulensis*, *E. globulus*, *E. urophylla*, *E. grandis* and *E. urophylla* x *E. grandis* hybrids (Mullins et al., 1997; Ho et al., 1998; Moralejo et al., 1998; Gonzalez et al., 2002; Tournier et al., 2003; Valerio et al., 2003). These are predominantly laboratory studies. Field studies of transgenic eucalypts have not been done in Australia, however, there have been reports of field trials in the UK and Spain in 1995, in Portugal and South Africa in 1997 and in Uruguay and Chile in 1997/1998 (Potts et al., 2001) including field testing of glyphosate resistant transgenic *E. grandis* (Llewellyn, 2000). With field studies of transgenic eucalypts comes the issue of genetic pollution, which can occur because of the high outcrossing rates of eucalypts and the propensity for interspecific hybridisation within the genus (Potts et al., 2003). Without the creation of sterile

transgenic plants, transgenes may be transferred through pollen from one individual or species to another (Barbour et al., 2003; Potts et al., 2003). Work has been done on creating sterile *E. occidentalis* at CSIRO Forestry and Forest Products (http://www.ffp.csiro.au/publicat/onwood/onwood39/story1.html).

Table 6. Research involving genetic modification in *Eucalyptus* species.

Species	Modified trait	Gene(s) altered	Reference
E. camaldulensis	Cellulose	*cbd, cel1*	Shani et al. (2003)
E. camaldulensis	Lignin	*C4H, CAD, Ntlim1*	Chen et al. (2001), Kawaoka et al. (2003), Valerio et al. (2003)
E. camaldulensis	Stress resistance	*DREB1A*	Hibino et al. (2002), Kondo et al. (2002, 2003)
E. camaldulensis	Salt stress tolerance	*cod A*	Yamada-Watanabe et al. (2003)
E. camaldulensis	Insect/herbicide resistance	*cry3A, bar*	Harcourt et al. (2000)
E. grandis	Cellulose	*cbd, cel1*	Shani et al. (2003)
E grandis x *E. urophylla*	Lignin	*CAD*	Tournier et al. (2003)
E. urophylla	Bacterial wilt resistance	*cecropin D*	Shao et al. (2002)

Sequence variation

The comparison of DNA sequences between individuals and species is useful for determining evolutionary relationships and for identifying those regions of the genome that are under functional constraint and those which have more plasticity. The transferability of DNA markers between species (Byrne et al., 1996; Marques et al., 2002) indicates that the primer sequences for amplification of these markers are conserved, however, these are only short stretches of DNA and do not imply a great deal about the general conservation of the genome sequence. Variation in the sequence of larger DNA stretches, particularly from the chloroplast genome and the nuclear encoded ribosomal DNA, have been examined extensively to study phylogeny, phylogeography and hybridisation in eucalypts. Variation and conservation of other nuclear genes, on the other hand, has been examined little.

DNA sequence variation in eucalypts was first studied looking at the chloroplast DNA (Steane et al., 1991). Since then chloroplast DNA has been used in many eucalypt studies aimed at phylogenetic and phylogeographic reconstruction and identification of patterns of hybridisation, with a large body of work developed based on the use of RFLP (Byrne et al., 1993; Byrne and Moran, 1994; Sale et al., 1996; Steane et al., 1998; Jackson et al., 1999; McKinnon et al., 1999; Byrne and

Macdonald, 2000; Byrne and Hines, 2004) and through sequencing variable regions such as JLA (Vaillancourt and Jackson, 2000; McKinnon et al., 2001a; McKinnon et al., 2001b; Stokoe et al., 2001) or JLA+ (Freeman et al., 2001; Whittock et al., 2003; McKinnon et al., 2004). In comparison to the chloroplast DNA, Vaillancourt et al. (2004) suggested the mitochondrial DNA would likely harbour little variability. Sequence variation in the nuclear encoded ribosomal DNA has also been examined for phylogenetic reconstruction in eucalypts (Udovicic et al., 1995; Steane et al., 1999; Steane et al., 2002). This has included the 5S ribosomal DNA from which the tandemly repeated genes were identified as highly conserved and the noncoding intragenic spacers highly variable in two *Angophora* species and 19 *Eucalyptus* species (including *Corymbia*) (Udovicic et al., 1995). The internal transcribed spacer (ITS) of the ribosomal DNA has also been examined and found to be highly informative for phylogenetic reconstruction in a large study sampling ninety species of *Eucalyptus*, six of *Angophora* and sixteen of *Corymbia* (Steane et al., 2002).

Few studies into the sequence variation of other nuclear genes have been undertaken in eucalypts. The first of these studies was conducted by Poke et al. (2003) looking at sequence variation in two lignin biosynthesis genes within 23 *E. globulus* individuals from two provenances (see Appendix 1). *CCR* was found to be three times more polymorphic than *CAD2* when comparing the number of SNPs per base pairs in exons of these genes (Poke et al., 2003). More of the SNPs in *CCR* were found to change the amino acid encoded, suggesting that *CAD2* may have more functional constraints. The introns of *CCR* were found to be more polymorphic and have greater incidence of insertions/deletions in comparison to exons (Poke et al., 2003). This study highlighted the variation that could be found at the intraspecific level in nuclear genes of eucalypts and how this was dependent on the gene and the part of the gene being examined.

What is next for *Eucalyptus*?

A vast amount of knowledge has already been gained from genomic research in eucalypts. Large segregating families have been established, high-resolution genetic maps constructed, large EST projects are coming close to fruition, and the sequencing of the chloroplast genome in *E. globulus* is complete. The sequencing of

the genome of a *E. camaldulensis* clone at a 4X shotgun coverage is now underway at the Kazusa DNA Research Institute in Japan and is expected to provide a draft sequence by mid-2007 (T. Hibino, personal communication). This will provide a valuable resource for eucalypt geneticists and will open the way for collaborative efforts by the eucalypt community including the creation of genetic linkage maps involving the *E. camaldulensis* clone. These maps may be aligned to linkage maps of other *Symphyomyrtus* species allowing the extrapolation of gene and QTL data to the genome sequence and will also aid in physical mapping, giving a significant boost to genomic research in *Eucalyptus*. The International *Eucalyptus* Genome Consortium (http://www.ieugc.up.ac.za) was established after a meeting in Hobart, Australia in 2004 for the purpose of uniting the community working in *Eucalyptus* genome research and provides a useful network and organisation to underpin these collaborative efforts.

The knowledge to be gained from a genome sequencing project for answering evolutionary, developmental and diversity questions is enormous. Genes in the regions underlying QTL can be identified, gene function characterised by identifying orthologs already found in *Arabidopsis*, and regulatory elements discovered (Brunner et al., 2004). As already realised with poplar, a perennial, woody species, it will allow for comparative genomics with annual crop plants (Stirling et al., 2003) to highlight developmental and evolutionary processes (Brunner et al., 2004; Tuskan et al., 2004). The sequencing of a second tree genome may also allow identification of parts of the genome which are associated with the tree habit in angiosperms. Comparison of the sequenced genomes with ESTs of other species with different characteristics (drought/cold/salt tolerance or insect resistance) may pinpoint genes which allow adaptation to harsh environments (Tuskan et al., 2004). It has been proposed that sequencing of a whole genome will also eventuate in the analysis of the entire set of proteins and metabolites (Tuskan et al., 2004). The sequencing of the *E. camaldulensis* genome and the formation of the International *Eucalyptus* Genome Consortium will open the way for efficient research efforts through integration of diverse projects and collaboration, providing a framework for future research in *Eucalyptus*.

CHAPTER 2

Predicting extractives and lignin contents in *Eucalyptus globulus* using near infrared reflectance analysis on ground wood

This chapter has been published as: Poke F.S., Wright J.K., Raymond C.A. (2004) Predicting extractives and lignin contents in *Eucalyptus globulus* using near infrared reflectance analysis. Journal of Wood Chemistry and Technology 24(1): 55-67.

Introduction

Eucalyptus globulus Labill. is one of the most important species used in temperate plantation forestry world wide for the production of pulp for paper. Its growth and wood properties make it ideally suited for this purpose. The wood is characterised by a high cellulose content and low extractives and lignin levels, which are conducive to high pulp yield (Miranda and Pereira, 2001). Although the wood properties of *E. globulus* are generally favourable, much variation has been found in density, pulp yield, extractives content, lignin content, and lignin composition between provenances (Dutkowski and Potts, 1999; Rodrigues et al., 1999; Muneri and Raymond, 2000; Miranda and Pereira, 2001; Raymond et al., 2001a).

One of the key economic drivers for kraft pulping is the yield of pulp per unit input of wood. Two factors influencing pulp yield are extractives and lignin contents of the wood, both of which are negatively correlated with pulp yield in *E. globulus* (Wallis et al., 1996; Miranda and Pereira, 2001). During kraft pulping, extractives and lignin must be chemically separated from the cellulose fibres to produce high quality pulp. Lignin is often the main component of wood after cellulose, comprising up to 36% of the dry weight. Extractives refer to a wide range of non-structural compounds within the plant, which are mostly soluble in organic solvents

or water. They are largely resin acids, fatty acids and their esters, waxes and unsaponifiable substances and some of the more polar compounds such as phenols, tannins, lignans, and low molecular mass carbohydrates (Appita Standard, 1994).

If considerable genetic variation exists within a species for pulp yield it is possible to breed for improvement. However, in tree breeding terms, pulp yield is a composite trait. For a given set of pulping conditions, the yield will be affected by differences in cellulose, hemicellulose, extractives and lignin contents plus lignin composition. In order to be able to understand the detailed genetic control of pulp yield, or to be able to manipulate the genes involved, it is essential to break the composite trait down into its component parts and examine the genetic control and interactions of each of the component traits separately. For some of the components of pulp yield, such as lignin, the biosynthetic pathways are understood (Baucher et al., 1998), opening up the option to genetically modify such traits by direct selection of favourable genotypes.

To be able to breed for any trait, or to explore potential for direct genotype selection, it is essential to be able to screen large numbers of individual trees. Measuring pulp yield, extractives or lignin content by traditional chemistry is costly and time-consuming. Currently, the commercial cost of measuring kraft pulp yield is A$1000 per sample and requires the destruction of the sample tree, which is undesirable from a tree breeding perspective. Extractives and lignin contents can both be measured using small wood samples removed from standing trees. Wood samples are air-dried and ground to wood meal prior to removing the extractives using a methanol extraction procedure for six to eight hours, which gives close to the total extractives content of wood (Appita Standard, 1994). The lignin analysis in which the wood meal is first digested with 72% sulphuric acid and subsequently allowed to degrade in 3% sulphuric acid, takes one and a half days per sample to complete (Appita Standard, 1978). The commercial cost for determining extractives and lignin contents is currently around A$350 per sample. Taking all of this into consideration, using extractives or lignin contents as selection criteria in operational breeding programs would not be particularly viable.

The use of an indirect method for determining the extractives and lignin contents of wood, such as near infrared reflectance (NIR) analysis, could provide a cheaper and quicker alternative, which would allow the determination of these traits in a larger number of samples and their incorporation in breeding programs. NIR analysis has previously been used for the prediction of pulp yield, (Wright et al., 1990; Michell, 1995; Schimleck and Michell, 1998) cellulose content (Wright et al., 1990; Schimleck et al., 1999; Raymond and Schimleck, 2002) and recently for extractives content, Klason lignin content and lignin composition (ratio of syringyl to guaiacyl sub-units) in *Eucalyptus* (Baillères et al., 2002). NIR analysis involves measuring the spectra of wood in the near infrared region, where a series of complex, overlapping bands are produced from the vibrations of chemical bonds of the components, which change with differences in the wood chemistry (Raymond and Schimleck, 2002). The region between 1200 and 2500 nm is the most useful of the NIR spectrum. Below 1200 nm measurements are difficult due to the weak absorption bands and above 2500 nm the strength of the bands is too great (Norris, 1989). NIR is ideal for screening large numbers of samples, as only a small quantity of wood is necessary (approximately 3 g air-dry) and spectral measurements can be taken relatively quickly. Analysis involves the development of a calibration using a large number of samples, where the NIR spectrum of the wood is related to the known extractives or lignin contents. Based on the NIR spectra, the extractives and lignin contents can be predicted for further samples, and the calibration validated using a "blind test". It is essential that the calibration represents the full scope of variation in the samples for the trait being analysed and that its development is not influenced by site variation. NIR analysis would, therefore, seem to be an ideal technique for determining extractives and lignin contents in a large number of wood samples.

The aim of this study was to develop calibrations for predicting extractives content, acid-soluble lignin content, Klason lignin content, and total lignin content in *E. globulus* using NIR analysis. Klason lignin is the proportion of the total lignin that is insoluble in acid. The NIR spectra were obtained from wood core samples from 155 trees at a single site, laboratory extractives and lignin content measurements were taken on a subset of these samples and calibrations developed. Calibrations were then used to predict the extractives content, acid-soluble lignin content,

Klason lignin content, and total lignin content of the remaining samples. The calibrations were validated using a "blind" test.

Materials and Methods

Field sampling

The wood samples of *Eucalyptus globulus* used in this study were obtained from a field trial located at West Ridgley, Tasmania (Gunns Ltd) based on the CSIRO Australian Tree Seed Centre collection and consisted of open-pollinated families (Gardiner and Crawford, 1987, 1988). A total of 155 trees were sampled from eight subraces of *E. globulus*, 37 families and at least two individuals per family (Table 1). Bark-to-bark wood cores were taken at breast height according to the method described by Raymond et al. (2001b). For each core only the outer quarter was used for analysis due to the presence of decay in the heartwood. These partial cores were used for NIR analysis and determination of extractives and lignin contents.

Table 1. *Eucalyptus globulus* samples used to develop NIR calibrations.

Subrace	Number of families	Number of individuals
Flinders Island	4	16
King Island	5	19
North-eastern Tasmania	4	16
South-eastern Tasmania	7	23
Southern Tasmania	4	15
Strzelecki Foothills	4	20
Strzelecki Ranges	4	21
Western Otways	5	25
Total	37	155

Determination of extractives and lignin contents

Wood cores were air-dried and cut into small pieces, further fragmented using an ESSA 200 mm disc pulverizer and ground to wood meal using a Wiley Mill with a 1 mm screen for 3 minutes. The extractives contents were determined according to the Appita Standard (1994). 2 g of wood meal was extracted with methanol in a Soxhlet apparatus for 6 hours. The solvent was recovered and the extractives

content determined from the mass of residue after drying at 105°C for one hour and reported as a percentage of original wood sample. Klason and acid-soluble lignin contents were determined on the extracted wood meal according to the Appita Standard (1978). 15 mL of 72% H_2SO_4 was added to 1 g of extracted wood meal and the mixture was placed in a shaking water bath at 20°C for 2 hours. The sample was diluted to a final concentration of 3% H_2SO_4 with boiling purified water and refluxed for 4 hours. The sample was vacuum filtered through a tared alundum crucible and washed with boiling purified water. Acid-soluble lignin was determined on the combined filtrate and washings by measuring absorbance at 206 nm using a UV/VIS spectrophotometer. Klason lignin was determined by the mass of residue after drying at 105°C overnight. Measurements were reported as a percentage of the original wood sample and combined to give the total lignin content. A reference sample was used to check the repeatability of the method throughout the course of the measurements.

Development of NIR calibrations

The NIR spectra were measured on the wood meal from each core in diffuse reflectance mode in a scanning spectrophotometer (NIRSystems Inc. Model 5000) at CSIRO (Clayton, Victoria). A ceramic standard was used as the instrument reference. Spectra were collected at 2 nm intervals over the 1100-2500 nm wavelength range. Fifty scans were performed per sample and the results averaged. The instrument's Vision® software was used to convert the data to the second-derivative mode using a segment width of 10 nm and a gap width of 0 nm.

Laboratory lignin contents were determined for a subset of 61 samples for the development of calibrations for predicting acid-soluble lignin, Klason lignin and total lignin contents. These samples were selected on the NIR predicted lignin content which was based on lignin results of a subset of 19 samples using two families of King Island and Strzelecki (Poke, 2001). The 61 samples were selected to cover the entire range of predicted lignin contents and to include at least one individual per family. Laboratory extractives contents were determined for 54 of these samples for calibration development. Extractives content and acid-soluble

lignin, Klason lignin and total lignin contents were then predicted for the remainder of the 155 samples.

A "blind" test was used to validate the calibrations. Six samples were selected by an independent person to cover the range of predicted lignin contents. Wood meal for each of these samples was transferred to new bottles, which were renumbered so that the identity of the samples was concealed. These six samples were then assayed in the laboratory and the results collated with the predicted values to determine regression relationships between predicted and observed values.

All calibrations were developed at CSIRO (Clayton, Victoria) using Vision® software (version 2.51) and partial least squares (PLS) analysis with four cross validation segments and a maximum of 10 factors (vectors). For cross validation, the calibration set was randomly split into four equal-size segments. One segment was excluded and the remaining segments were used to develop PLS calibrations with different numbers of factors. The calibrations were checked to determine how well they fitted the samples in the excluded segment. The process was repeated four times until all samples were used and validated.

The number of factors used in the final calibration is very important as too many factors will overfit the data and give a calibration that may only be suited to the data from which it is was derived. If not enough factors are selected, variation in the data will not be sufficiently described and under fitting occurs (Martens and Naes, 1991). The optimum number of factors for a calibration can be identified when the mean square error of cross validation (MSECV) reaches a minimum (Miller, 1989). Each calibration also has an R-value, which is the correlation coefficient from the PLS regression.

The degree of fit of the calibration to the data is given by the standard error of calibration (SEC), where $\hat{y}_i$ is the chemical content for validation sample i estimated using the calibration, y_i is the known chemical content of sample i, NC is the number of samples used to develop the calibration, and k is the number of factors used to develop the calibration (Miller, 1989; NIR Systems Inc., 1990; Workman, 1992).

$$SEC = \sqrt{\frac{\sum_{i=1}^{NC}(\hat{y}_i - y_i)^2}{(NC - k - 1)}}$$

Results

Repeatability of extractives and lignin content determinations

Determination of the extractives content of a reference sample four times throughout the course of assays showed laboratory measurements to be highly repeatable with a standard deviation of 0.18% (Table 2). The lignin content of the same reference sample was measured seven times throughout the course of the lignin assays and gave repeatable results for the total lignin content with a standard deviation of 0.46% (Table 2).

Table 2. Statistics for extractives content and total lignin content of a reference sample measured multiple times over the period of assays.

Reference sample	Mean (%)	Standard deviation	Minimum (%)	Maximum (%)
Extractives content	3.22	0.18	3.03	3.37
Total lignin content	27.43	0.46	26.89	28.21

Prediction of extractives and lignin contents

A high R^2 (correlation coefficient from PLS regression) value of 0.93 and a SEC of 0.56 were obtained for the calibration for extractives content using measurements from 54 samples (Table 3). Calibrations developed for both Klason and total lignin contents using lignin measurements for 61 samples had high R^2 values of 0.93 and 0.88, respectively, and SECs of 0.40 and 0.54 (Table 3). The calibration developed for acid-soluble lignin had a moderate R^2 value of 0.62 and SEC of 0.39 (Table 3). Based on the predicted lignin contents from these calibrations, a subset of six samples was used as a "blind test". Predicted and laboratory extractives contents

were highly correlated with an R^2 value of 0.89 (Figure 1). For acid-soluble lignin contents the laboratory and predicted values were also well correlated with an R^2 of 0.83 (Figure 2). Predicted and laboratory Klason lignin and total lignin contents were very highly correlated with an R^2 of 0.97 for Klason lignin content and 0.99 for total lignin content (Figures 3 and 4).

Table 3. Statistics for the NIR calibrations for extractives content, acid-soluble lignin content, Klason lignin content and total lignin content in *Eucalyptus globulus*.

Wood component	No. of Factors	R^2	SEC
Extractives content	5	0.93	0.56
Acid-soluble lignin content	7	0.62	0.39
Klason lignin content	6	0.93	0.40
Total lignin content	6	0.88	0.54

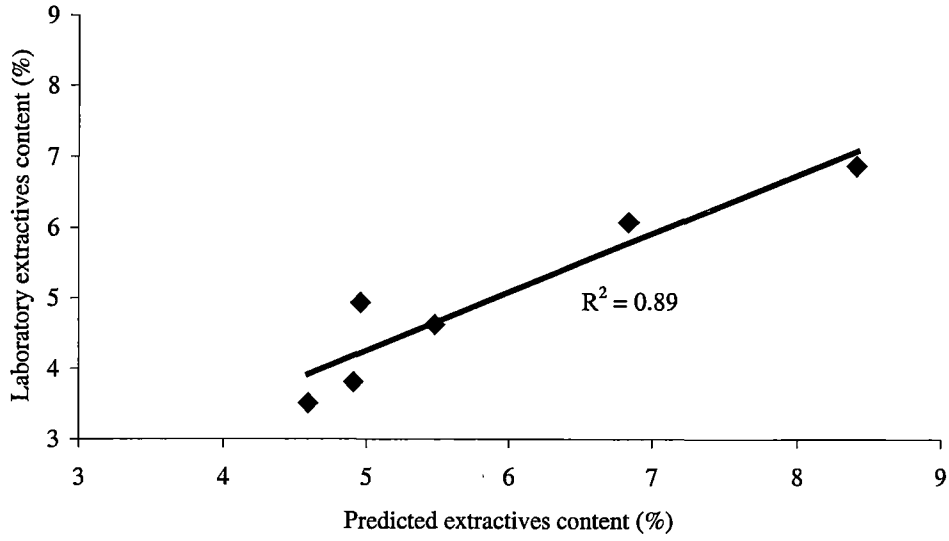

Figure 1. Relationship between NIR predicted and laboratory extractives content for six samples of *Eucalyptus globulus* used in a blind test of the calibration.

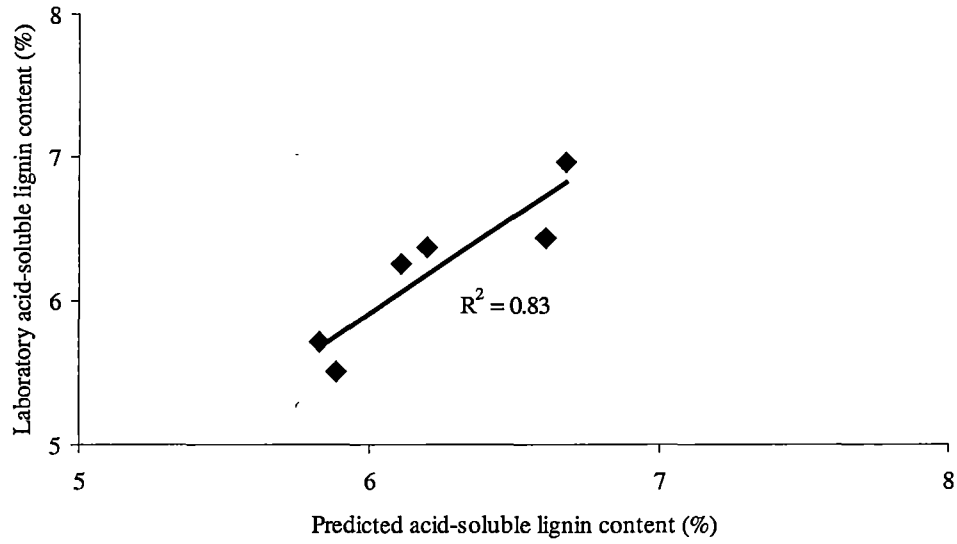

Figure 2. Relationship between NIR predicted and laboratory acid-soluble lignin content for six samples of *Eucalyptus globulus* used in a blind test of the calibrations.

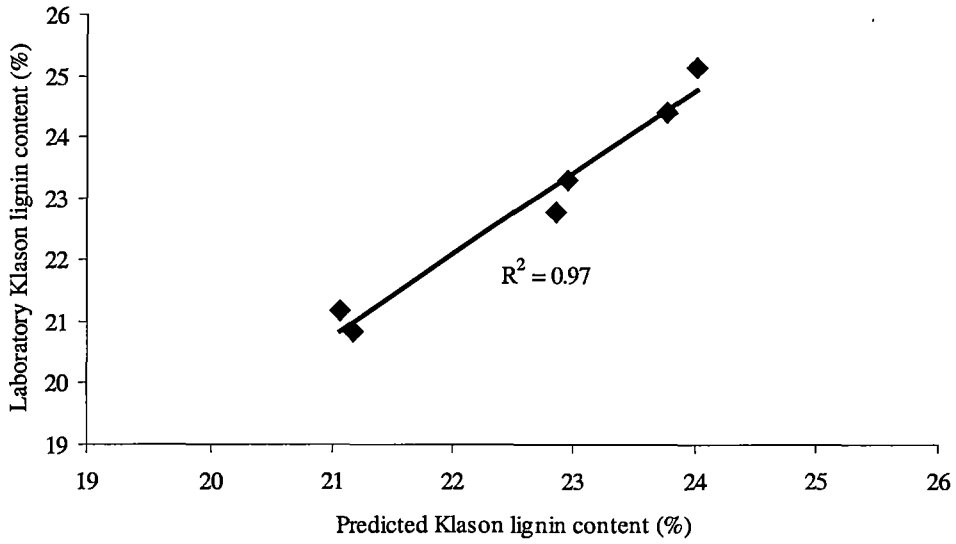

Figure 3. Relationship between NIR predicted and laboratory Klason lignin content for six samples of *Eucalyptus globulus* used in a blind test of the calibrations.

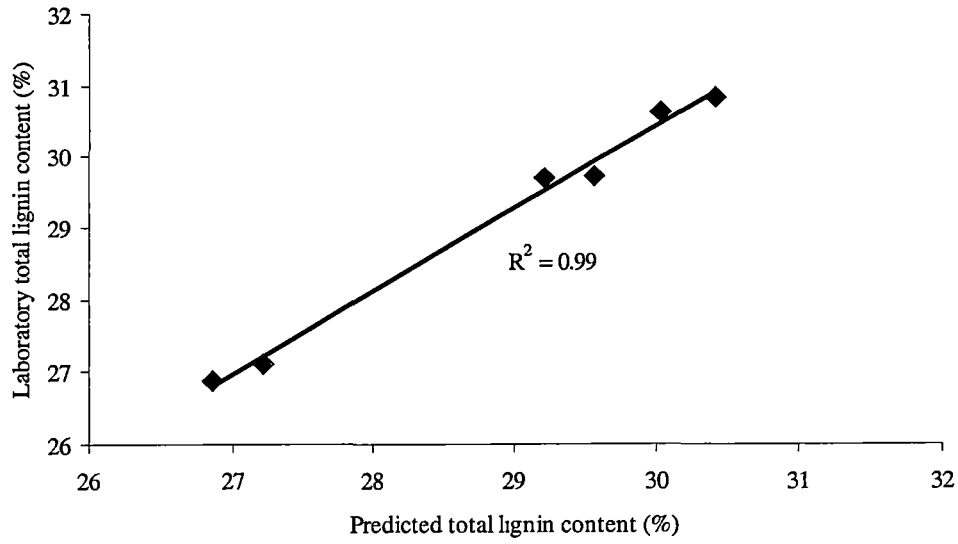

Figure 4. Relationship between NIR predicted and laboratory total lignin content for six samples of *Eucalyptus globulus* used in a blind test of the calibrations.

Discussion

Good calibrations were developed for predicting extractives and lignin contents in *E. globulus* using NIR analysis, providing a cost-effective method for screening large numbers of wood samples for these traits. These calibrations were used to predict the extractives content of 101 samples and the acid-soluble lignin, Klason lignin and total lignin contents of 94 samples. SEC were less than 0.6% and high correlations were obtained between laboratory and predicted values in the "blind" test (R^2 of 0.89 for extractives content, 0.83 for acid-soluble lignin content, 0.97 for Klason lignin content and 0.99 for total lignin content), indicating NIR analysis can be reliably used to predict extractives and lignin contents in wood samples. The moderate correlation coefficient obtained for the calibration of acid-soluble lignin content (0.62) may be due to this trait having small values and less variation within these values, and may also suggests that lignin composition is having an influence on the acid-soluble lignin values. Baillères et al. (2002) in a study of progeny from a *E. urophylla* x *E. grandis* cross obtained a slightly lower R^2 value for the calibration of the extractives content, and found a weaker correlation between predicted and laboratory values than that determined here. Baillères et al. (2002) also obtained a similar R^2 value for the calibration of the Klason lignin content

(0.87), however, predicted and laboratory values showed a much lower correlation of 0.83 than that determined here. The similar results between this study and the one of Baillères et al. (2002) suggest that NIR analysis would be an effective method for determining extractives and lignin contents across a range of *Eucalyptus* species.

The use of the NIR spectra of wood samples to predict extractives and lignin contents has important implications for using these traits as selection criteria in tree breeding programs. Currently, the focus lies in the selection of trees based on basic density and growth, predominantly due to the cost effectiveness and ease of measuring these traits. Pulp yield, the key trait for selection, can only be measured using a destructive sampling technique and is costly to measure. In addition, pulp yield is influenced by all the components of the wood, and therefore, in order to select trees in breeding programs based on genotype, pulp yield must be broken down into its composite traits. The negative impact of both extractives and lignin contents on the pulping process, together with the good characterisation of lignin biosynthesis within plants, suggests that these wood traits may be good candidates for use in improving yields and ultimately achieving economic gains in the industry. The traditional commercial method for measuring these traits is a costly and time-consuming process with a low sample throughput. The speed and low cost associated with using NIR makes it a very attractive alternative for predicting extractives and lignin contents in wood samples over traditional chemistry means, and allows for the processing of large numbers of samples using only a small amount of wood from an increment core. Therefore, screening large numbers of trees for these traits for use in breeding programs is a viable option. The ability to predict the individual components of the lignin content also has important implications and can assist with selecting more favourable trees for pulping.

The calibrations developed in this study were based on *E. globulus* samples from a single site. The real advantage of using NIR analysis to predict extractives and lignin contents on the large scale necessary in tree breeding programs will become apparent with the development of multi-site calibrations. Variation in these traits may arise through differences in growing conditions and also through the variation in genetic makeup within a single species. Global calibrations will be necessary for using this technique for screening a wide range of trees, so the appropriate

calibration can be used for new samples. Development of global calibrations requires assessment of the NIR spectra and wet chemistry analysis for trees from a range of sites. Spectra for trees from new sites can then be evaluated to determine whether they fit in the existing calibration model space, or fall outside. Where new samples fall outside of the model space, the spectra, together with associated wet chemistry data, can be used to expand the calibration space. In this way broader global calibrations can be built up over time. As further sites are included, the requirement for wet chemistry analysis is reduced as it is only required on a subset of samples to ensure that the predictions are correct. Development of such global calibrations will allow both extractives and lignin content determination in *E. globulus* from a wide range of locations, where large variations in these traits are likely to be found and can be exploited in breeding programs.

Since the contents of this chapter was published in Poke et al. (2004), two other studies have been published on this topic and support the findings here. Good NIR calibrations have been developed for Klason lignin content using five species of *Pinus*, with correlation coefficients greater than 0.89 and SEC less than 0.42 (Hodge and Woodbridge, 2004). Strong correlations between predicted and laboratory values (R^2 between 0.77 and 0.91) were also obtained for a separate set of samples (Hodge and Woodbridge, 2004). Calibration statistics were comparable to what was found for *E. globulus*, although correlations between predicted and laboratory values were lower for *Pinus*. NIR spectroscopy has also been used to develop a calibration for the prediction of total lignin content in *E. camaldulensis* (Terdwongworakul et al., 2005). A correlation coefficient of 0.92 was observed for this calibration, which was slightly higher than that for *E. globulus*. However, the correlation between predicted and laboratory values for a separate set of samples (R^2 of 0.88) was much lower than that obtained for *E. globulus*. NIR spectroscopy appears to be a reliable predictor of lignin content for a range of genera and also for different species of *Eucalyptus*.

CHAPTER 3

Predicting extractives, lignin and cellulose contents using near infrared spectroscopy on solid wood in *Eucalyptus globulus*

This chapter has been accepted for publication as: Poke F.S., Raymond C.A. (2006) Predicting extractives, lignin and cellulose contents using near infrared spectroscopy on solid wood in *Eucalyptus globulus*. Journal of Wood Chemistry and Technology 26(2): in press.

Introduction

Eucalyptus is grown throughout the world for pulp and solid wood products. In temperate regions, *Eucalyptus globulus* is one of the major species grown for pulp production. Its wood is ideally suited to this purpose, with a high cellulose content, and low lignin and extractives contents, which contribute to a high pulp yield (Miranda and Pereira, 2001). Breeding programs have been established for this species with the objective of further improving some of these characteristics. However, these programs require assessing large numbers of trees for multiple traits. For this purpose, techniques have been developed for increasing the efficiency of this process. Near infrared reflectance (NIR) spectroscopy has been shown to be a reliable predictor of wood chemistry from increment cores in *Eucalyptus* (Wright et al., 1990; Schimleck and Michell, 1998; Schimleck et al., 1999; Raymond et al., 2001; Baillères et al., 2002; Raymond and Schimleck, 2002; Poke et al., 2004; Terdwongworakul et al., 2005; Chapter 2). NIR calibrations developed for the prediction of pulp yield, cellulose content, lignin content, lignin composition and extractives content have all been successful in predicting these traits in larger data sets with considerable accuracy. This technique has provided an excellent alternative to using traditional wet chemistry methods, which are costly

and time-consuming, and has become a well used tool for the rapid and cost effective prediction of important chemical wood traits in tree breeding programs.

Tree breeding programs require large numbers of trees to be screened in the most rapid and cost-effective way. NIR analysis has certainly helped with this, making it possible to predict a number of traits from a single increment core. Improvements can still be made, however, to increase the efficiency of using NIR analysis as a prediction tool for wood chemistry. Existing calibrations for *Eucalyptus* (mentioned above) were developed using NIR spectroscopy on ground wood, and this processing requirement is in itself quite costly and time-consuming. Increment cores must be ground to wood meal using a Wiley mill with a one mm screen, for a total of three minutes per sample (Poke et al., 2004; Chapter 2). The mill must also be cleaned between samples, therefore making it possible to process only 12 samples per hour. This grinding step can become quite costly when looking at sampling large field trials. Using NIR spectroscopy on solid wood could allow the removal of this step from wood chemistry evaluation, increasing the ease, cost-effectiveness and rapidity of NIR as a prediction tool.

It is unknown if current ground wood calibrations can be used to predict traits based on spectral data from solid wood. There are also very few studies into the use of NIR analysis on solid wood for calibration development and trait prediction. The reason for this may involve the limitation of using reflectance measurements on solid wood, which primarily includes the small penetration depth (1-4 mm) into the sample (Yeh et al., 2004). For nonhomogenous samples such as wood this may lead to large variation in the results, and a strong dependence on sample size and preparation technique (Yeh et al., 2004). Despite this potential limitation, NIR analysis has previously been used on solid wood to develop calibrations for predicting physical traits such as density, stiffness (longitudinal modulus of elasticity), strength (modulus of rupture) and microfibril angle in *Eucalyptus delegatensis*, *Pinus radiata* and *Larix decidua*, giving excellent results for both hardwoods and softwoods (Gindl et al., 2001; Schimleck et al., 2001; Schimleck et al., 2002). However, the use of NIR analysis for the prediction of the chemical composition of solid wood is a relatively new area to be examined. Recently, this has been investigated in loblolly pine (*Pinus taeda*) with encouraging results

(Kelley et al., 2004). Kelley et al. (2004) successfully developed calibrations and predicted values for lignin, extractives, glucose, xylose, mannose and galactose contents of solid wood. The calibrations for each trait had high correlation coefficients, and predicted and laboratory values were well correlated. To date this is the only published study detailing the prediction of chemical composition from solid wood using NIR, and so far this technique has not been examined in hardwoods such as *Eucalyptus*.

This study aimed to extend the findings of Chapter 2 (Poke et al., 2004) by exploring the possibility of using solid wood, to increase the rapidity and cost-effectiveness of using NIR as a prediction tool for wood chemistry in *Eucalyptus*. The use of current ground wood calibrations for the prediction of wood chemistry based on solid wood NIR spectral data, was evaluated in *E. globulus*. Further, new calibrations specifically for solid wood were developed for the prediction of the chemical wood traits. This would exclude the need for the grinding of increment core samples, and increase the speed and cost-effectiveness of screening samples for wood chemical composition using NIR analysis. NIR spectra were collected for pith-to-bark strips, and extractives, acid-soluble lignin, Klason lignin, total lignin and cellulose contents were predicted from current ground wood calibrations (Poke et al., 2004; Chapter 2). Laboratory measurements were made for extractives, acid-soluble lignin, Klason lignin, total lignin and cellulose contents on a subset of samples to evaluate these predictions. Further laboratory measurements were made for these traits to develop calibrations specifically for solid wood. Calibrations were used to predict these traits in the remainder of the data set, and were validated with further wet chemistry measurements.

Materials and Methods

Field sampling

Wood samples of *E. globulus* at 14 years of age were obtained from a base population field trial located at West Ridgley, Tasmania (Gunns Ltd). This trial was based on the CSIRO Australian Tree Seed Centre collection and is composed of open-pollinated families (Gardiner and Crawford, 1987, 1988). The trees sampled

were the same used for the study on within-tree variation in wood chemistry detailed in Chapter 4. Ten trees were selected in pairs from the same family, the same part of the trial (north or south), to span the range of NIR predicted lignin contents obtained using the calibrations described in Chapter 2 (Poke et al., 2004), and to be decay free (noted in the absence of decay in wood cores previously taken in 2002). Trees near the edges of the trial were also excluded as they may be wind damaged and contain tension wood. Due to these restraints, trees were selected to have a lignin content range of 26 to 30%. At the time of sampling tree 10 was found to be forked close to the base, and therefore was excluded.

The nine trees were felled, total tree height was measured and disks 15 cm deep were cut at each 10% increment of total tree height (or as close as possible to avoid branches and defects). The base represented 0% and disks were taken up to 70% of tree height, after which the stem generally became too small. From the disks, 20 mm wide by 15 cm deep bark-to-bark strips were cut through the pith in the same orientation for each tree using a band saw. This segment was then cut in half down the pith, and a 20 mm deep pith-to-bark strip was taken either from the bottom or top (to avoid defects) of a constant side for each disk. The 20 x 20 mm strip was air-dried and further used for NIR analysis and wood chemistry measurements.

NIR analysis

The NIR spectra were taken in increments along the transverse face of each pith-to-bark strip at CSIRO (Clayton, Victoria). The first increment was 40 mm in length (longer so as the strip could be held in the machine) with the following increments 20 mm in length. Spectra were taken as described in Chapter 2 and Poke et al. (2004).

Evaluation of wood chemistry predictions using ground wood calibrations

Using the ground wood calibrations developed in Chapter 2 (Poke et al., 2004) the extractives, acid-soluble lignin, Klason lignin and total lignin contents were predicted for each increment. Cellulose content was also predicted using a CSIRO calibration. Using these NIR prediction estimates, samples were selected for wood

chemistry analysis so as to cover the range of predicted values, and in the majority of cases, so they would occur at the end of a pith-to-bark strip for easy sampling. Increments were cut from the strips and ground to wood meal as described in Chapter 2 (Poke et al., 2004). Extractives and lignin contents were measured according to the Appita Standards (1978, 1994) on 13 and 14 samples, respectively, for evaluation of the predictions. Crude cellulose content was measured in duplicate on nine samples for calibration validation using the diglyme method of Wallis et al. (1997). 10 mL of diglyme and 2 mL of concentrated hydrochloric acid were added to 1 g of wood meal in a 50 mL reaction bottle, which was sealed with a teflon coated cap, and placed in a shaking water bath at low speed at 90°C for 1 hour. The residue was collected in a tared alundum crucible by vacuum filtration, washed with 50 mL of methanol and followed by 250 mL of boiling water, and dried overnight at 105°C. Cellulose content was determined from the mass of residue after drying, and was reported as a percentage of original wood sample after adjustment for moisture content. Regression relationships were determined between predicted and laboratory values to evaluate the accuracy of the predictions

Solid wood calibration development and validation

Extractives and lignin contents were measured on a further 27 and 29 increment samples, respectively, for calibration development. Crude cellulose content was measured in duplicate on a further 30 samples for calibration development. Calibrations were developed for extractives, acid-soluble lignin, Klason lignin, total lignin and cellulose contents using the method detailed in Chapter 2 (Poke et al., 2004) at CSIRO (Clayton, Victoria). These calibrations were then used to predict the wood chemical composition of the remaining increments (253 increments for extractives content, 250 for lignin content and 254 for cellulose content). Seven to nine of these increment samples were then selected which spanned the range of predicted extractives, lignin and cellulose values, laboratory measurements were made, and regression relationships between predicted and laboratory values were determined.

Results

Evaluation of the NIR predictions from ground wood calibrations

Ground wood calibrations were unsuitable for predicting extractives and acid-soluble lignin contents from solid wood spectral data, with many of the predicted values being negative numbers (data not shown). Predictions of Klason lignin content for 14 samples were poorly correlated with laboratory values with an R^2 of 0.27 (Figure 1). Predicted total lignin contents for 14 samples were moderately correlated with laboratory measurements with an R^2 value of 0.54 (Figure 2). Predicted and laboratory values for cellulose content for nine samples were moderately well correlated with an R^2 value of 0.63 (Figure 3). Despite the moderate correlation coefficients for total lignin and cellulose contents (Figures 2 and 3), the regression equations indicated that the predictions lacked sufficient accuracy with X multipliers of 0.44 and 0.48, respectively (should be close to 1 for accurate predictions). Predictions for total lignin content were either under- or over-estimated (shown by the regression line crossing the 1:1 line), whereas cellulose content was generally overestimated. These regression lines were generally driven by outliers at one end of the extreme for both total lignin content and cellulose content (three and two points, respectively), and when these were removed multipliers on the X factor increased dramatically to 0.80 for total lignin content, and 0.69 for cellulose content, indicating that the majority of NIR predictions were quite close to the laboratory values (Figures 2 and 3).

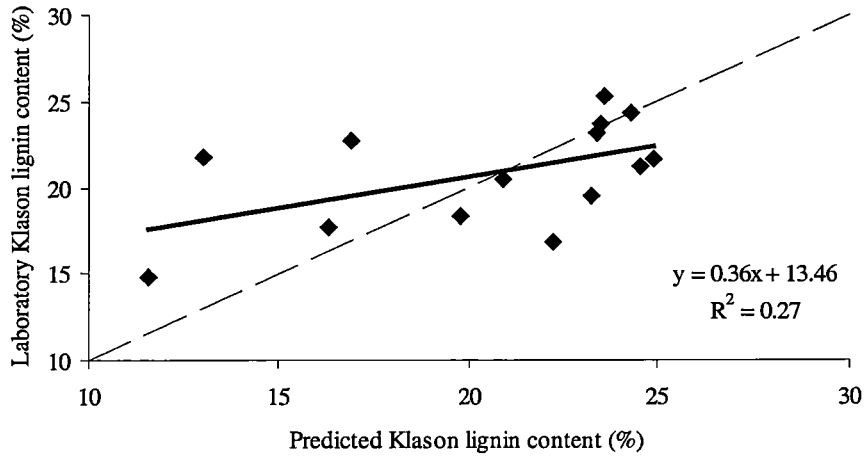

Figure 1. Relationship between NIR predicted and laboratory Klason lignin content (solid bold line) for 14 samples of *Eucalyptus globulus*. The 1:1 line is shown as a broken line. Predictions were made for solid wood using ground wood calibrations (Poke et al., 2004).

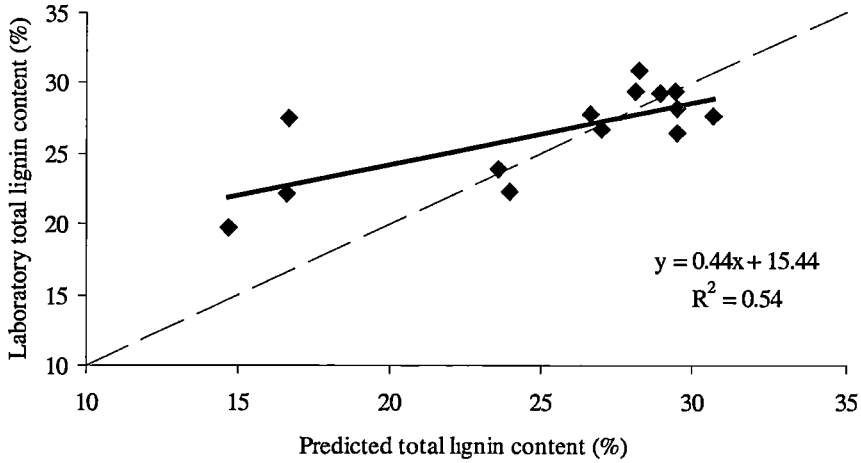

Figure 2. Relationship between NIR predicted and laboratory total lignin content (solid bold line) for 14 samples of *Eucalyptus globulus*. The 1:1 line is shown as a broken line. Predictions were made for solid wood using ground wood calibrations (Poke et al., 2004).

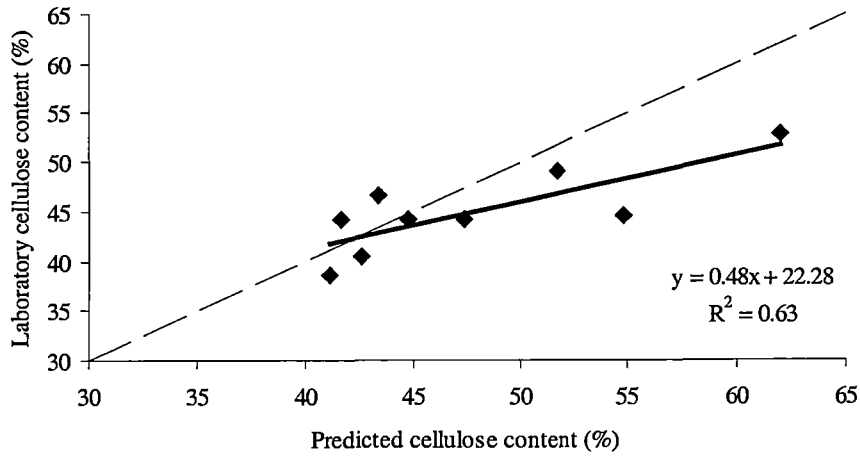

Figure 3. Relationship between NIR predicted and laboratory cellulose content (solid bold line) for nine samples of *Eucalyptus globulus*. The 1:1 line is shown as a broken line. Predictions were made for solid wood using ground wood calibrations (CSIRO, Clayton).

NIR calibrations for solid wood

Good calibrations were developed for predicting extractives, lignin and cellulose contents from solid wood (Table 1). A high R^2 value of 0.84 and a SEC of 1.37 were obtained for the calibration for extractives content using measurements from 40 samples (Table 1). Calibrations developed for acid-soluble lignin, Klason lignin and total lignin contents had R^2 values of 0.72, 0.78 and 0.76, respectively, and SECs of 0.41, 1.02 and 1.07 (Table 1) based on 43 samples. For cellulose content an R^2 value of 0.88 and a SEC of 1.14 were obtained for the calibration using 39 samples (Table 1). Based on the range of predicted extractives, lignin and cellulose contents, laboratory measurements were made for additional increment samples to validate the calibrations. Predicted and laboratory extractives contents for nine samples were highly correlated with an R^2 of 0.87 (Figure 4). Predictions were found to be quite accurate with the regression equation giving a multiplier on the X factor close to 1 (1.14). Generally predictions were either over- or under-estimated with the regression line crossing the 1:1 line (Figure 4). For acid-soluble lignin content a poor correlation was obtained between predicted and laboratory values for seven samples with an R^2 of 0.12 (Figure 5). Predicted and laboratory values for Klason lignin and total lignin contents for seven samples showed high correlations

with R^2 values of 0.79 and 0.67, respectively (Figures 6 and 7). Regression equations had multipliers on the X factor of 1.04 for Klason lignin content, and 1.40 for total lignin content, indicating the predictions were quite accurate, particularly for Klason lignin content. Predictions for Klason lignin content were consistently overestimated by half a unit (regression line lay parallel to the 1:1 line and was displaced to the right), whereas predictions for total lignin content were either over- or under-estimated (regression line crossed the 1:1 line) (Figures 6 and 7). Predicted and laboratory cellulose contents for nine samples were well correlated with an R^2 of 0.69 (Figure 8). The regression equation indicated that the predicted values were quite accurate with a multiplier on the X factor of 0.96. Predicted cellulose contents were generally overestimated by a half unit, as indicated by the regression line being parallel to the 1:1 line and displaced to the right (Figure 8). Two outliers were evident in this graph and when removed from the dataset the R^2 increased to 0.97.

Table 1. Statistics for the NIR calibrations for extractives, acid-soluble lignin, Klason lignin, total lignin and cellulose contents for solid wood in *Eucalyptus globulus*.

Wood component	Number of factors	R^2	SEC
Extractives content	6	0.84	1.37
Acid-soluble lignin content	6	0.72	0.41
Klason lignin content	4	0.78	1.02
Total lignin content	4	0.76	1.07
Cellulose content	5	0.88	1.14

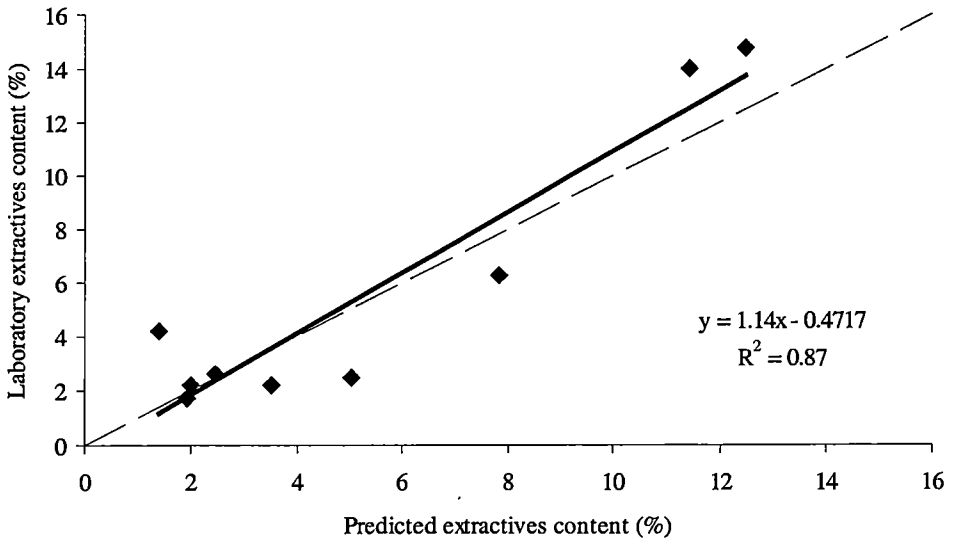

Figure 4. Relationship between NIR predicted and laboratory extractives content (solid bold line) for nine samples of *Eucalyptus globulus* used to validate the calibrations. The 1:1 line is shown as a broken line.

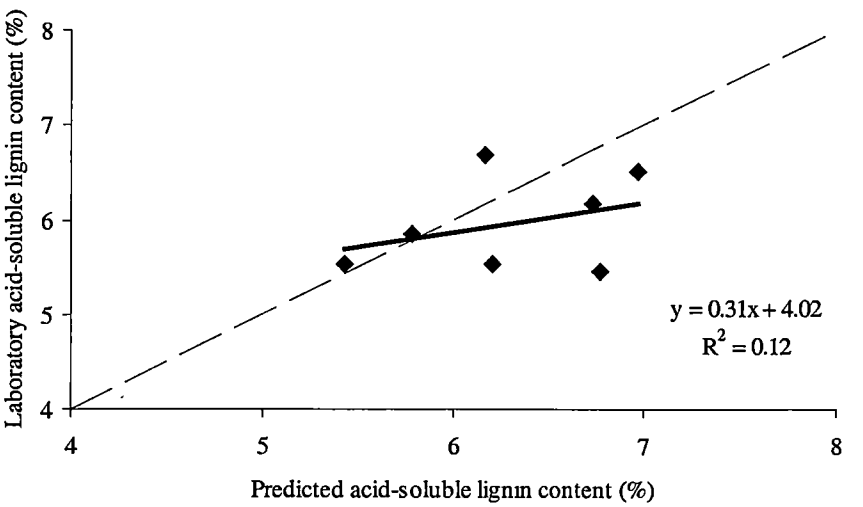

Figure 5. Relationship between NIR predicted and laboratory acid-soluble lignin content (solid bold line) for seven samples of *Eucalyptus globulus* used to validate the calibrations. The 1:1 line is shown as a broken line.

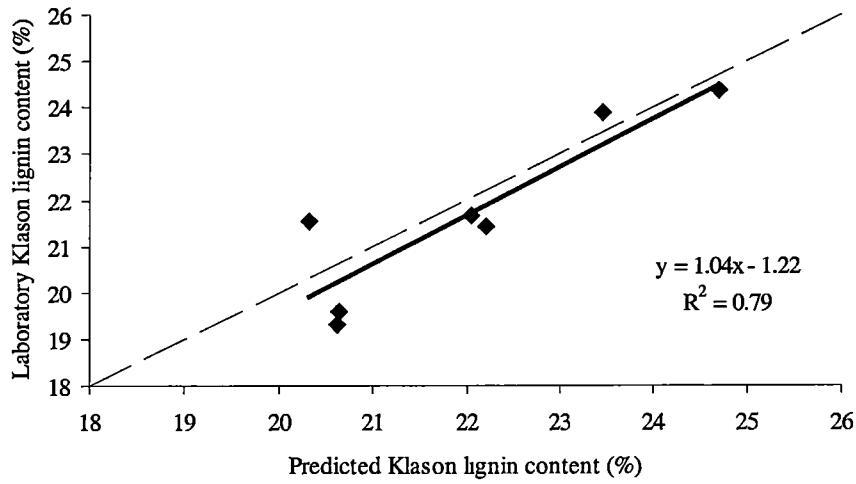

Figure 6. Relationship between NIR predicted and laboratory Klason lignin content (solid bold line) for seven samples of *Eucalyptus globulus* used to validate the calibrations. The 1:1 line is shown as a broken line.

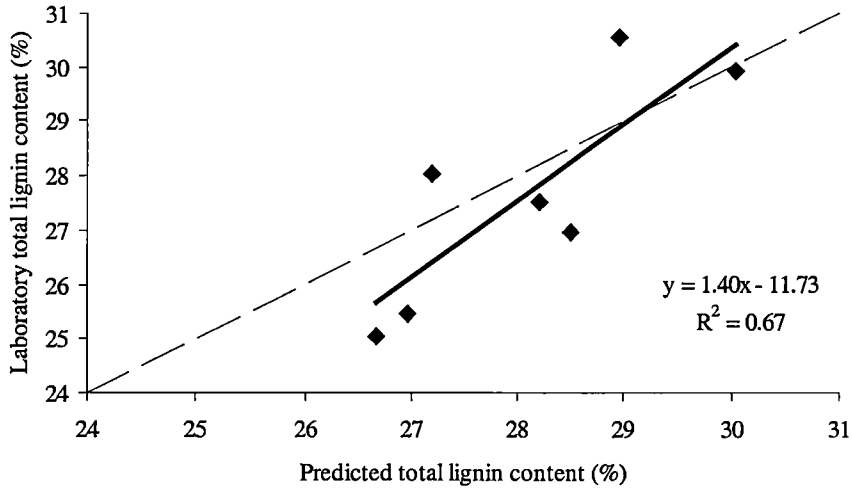

Figure 7. Relationship between NIR predicted and laboratory total lignin content (solid bold line) for seven samples of *Eucalyptus globulus* used to validate the calibrations. The 1:1 line is shown as a broken line.

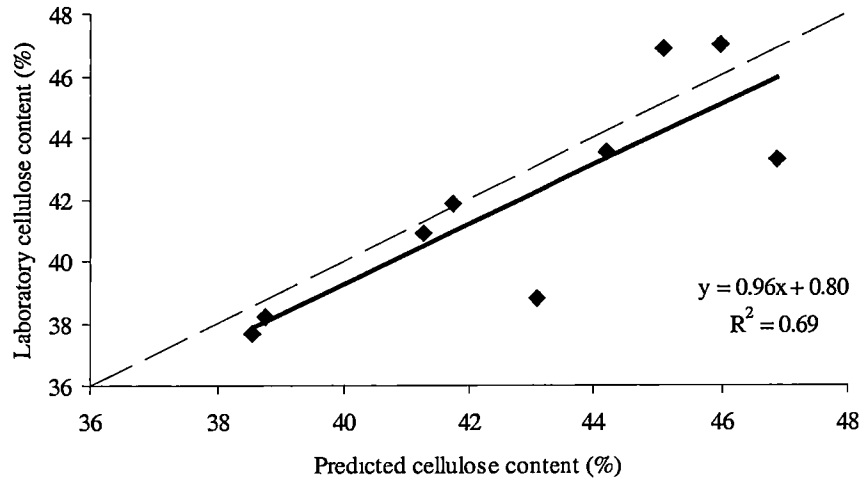

Figure 8. Relationship between NIR predicted and laboratory cellulose content (solid bold line) for nine samples of *Eucalyptus globulus* used to validate the calibrations. The 1:1 line is shown as a broken line.

Discussion

Existing ground wood calibrations (Poke et al., 2004; Chapter 2) were found to be unsuitable for the prediction of extractives, acid-soluble lignin and Klason lignin contents from solid wood. Moderate correlation coefficients were found between predicted and laboratory values for total lignin content (R^2 value of 0.54) and cellulose content (R^2 value of 0.63), suggesting this method could be used to give crude estimates of these traits, particularly cellulose content. Although regression equations indicated that these predictions did not have high accuracy overall, the majority of predictions were close to laboratory values and therefore ground wood calibrations may be useful for obtaining approximate values for these traits. Total lignin content was either under- or over-estimated, and cellulose content was generally overestimated. To obtain more accurate NIR predictions, calibrations for extractives, lignin and cellulose contents will need to be re-developed for solid wood. Ground wood calibrations may not work as well for solid wood because the spectra taken from ground wood involve the cell wall polymers being at varying angles incident to the NIR radiation within the one sample. For solid wood spectra the cell wall polymers will be at a consistent incident angle to the NIR radiation within the same sample. The spectra will differ between the two sample types as a result. Overall this study indicated that it is unlikely that existing wood chemistry

calibrations, that have been developed using ground wood, can be used for accurate NIR prediction of these traits from solid wood.

Good calibrations were developed using NIR analysis for predicting extractives, lignin and cellulose contents for solid wood in *E. globulus*. Calibration coefficients were generally high and above 0.72, and SEC were below 1.37%. Validation of these calibrations produced good correlations between NIR predicted and laboratory values for all traits except acid-soluble lignin content (R^2 of 0.87 for extractives content, 0.12 for acid-soluble lignin content, 0.79 for Klason lignin content, 0.67 for total lignin content and 0.69 for cellulose content). Klason lignin and cellulose contents were consistently overestimated by a half unit, whereas total lignin content was either over- or under-estimated slightly. These results suggest that calibrations developed using NIR spectra from solid wood, could be reliably used to predict extractives, lignin and cellulose contents with considerable accuracy. The good calibration for extractives content is particularly surprising as it is a composite trait comprised of a range of non-structural compounds that are soluble in organic solvents and water (Appita Standard, 1994). Therefore, prediction of extractives content from the NIR spectra, particularly from solid wood, would be expected to be more difficult, unlike pure compounds such as cellulose. The less successful prediction of acid-soluble lignin content may be due to this trait having small values with less variation between them, and it may be possible that lignin composition is having an effect on this trait, which has been suggested previously when NIR was used to predict this trait from ground wood (Poke et al., 2004; Chapter 2). Although acid-soluble lignin content could not be reliably predicted using this calibration, estimates for Klason lignin content and total lignin content were good, with Klason lignin content proving to be the most reliable prediction trait for lignin content from solid wood. Together these results suggest that calibrations developed for solid wood can be reliably used to predict wood chemistry, and will provide a rapid and cost-effective alternative to using ground wood calibrations.

The only published study detailing the development of NIR calibrations for chemistry prediction from solid wood is in loblolly pine (Kelley et al., 2004). Kelley et al. (2004) developed good calibrations with high correlation coefficients for both lignin (0.81) and extractives (0.93) contents, which are higher than those established in this current study. A stronger correlation between predicted and

laboratory values for lignin content was also obtained, with a correlation coefficient of 0.76, and a similar correlation coefficient was found for extractives content of 0.85, which was slightly lower than that obtained here (Kelley et al., 2004). The variation in results between these studies are likely to be due to differences between the genera used, *Eucalyptus* and *Pinus*, and to differences in the type of wood being used, softwood versus hardwood. However, results from both studies suggest that NIR scans of solid wood for calibration development and trait prediction will be a viable option. Axrup et al. (2000) have also shown that NIR analysis can be used to predict the extractives and Klason lignin contents of wood chips moving along a conveyor belt. However, the root mean square errors of calibration and prediction were consistently higher than those reported by Kelley et al. (2004), suggesting this technique is not as reliable.

It has been suggested that the quality of NIR calibrations developed for solid wood may be reduced (Kelley et al., 2004), which has also been noted here. Calibrations developed for ground wood of *E. globulus* (Poke et al., 2004; Chapter 2) had very high correlation coefficients compared to those established using solid wood. The accuracy of the predictions for ground wood were also excellent with regression relationships giving very high R^2 values for Klason lignin and total lignin contents (0.97 and 0.99, respectively), and high R^2 values for acid-soluble lignin and extractives contents (0.83 and 0.89, respectively). For cellulose content, Raymond and Schimleck (2002) developed a range of ground wood calibrations for different sites of *E. globulus*, and obtained moderate to very high correlation coefficients ranging from 0.78 to 0.94. Predicted and laboratory values were also highly correlated with R^2 values between 0.82 and 0.95 (Raymond and Schimleck, 2002). The values associated with ground wood calibrations are generally higher than those obtained for solid wood, however, the advantages to using solid wood calibrations may outweigh the slight decrease in accuracy for many applications. The increase in the ease and speed, and the decrease in the cost associated with removal of the wood grinding step, is a major advantage for using NIR analysis on solid wood for wood chemistry prediction, in preference to ground wood. Solid wood NIR calibrations can provide a useful and cost-effective tool for the rapid screening of large numbers of trees for wood chemical composition in breeding programs.

CHAPTER 4

Within-tree variation in lignin, cellulose and extractives in *Eucalyptus globulus*

Introduction

Traditionally, measuring trees for wood quality traits required the destructive sampling of the entire tree, which was costly, time-consuming and did not allow further use of that tree for breeding (Raymond et al., 1998). More recently, non-destructive techniques have been developed including the use of increment wood cores (Raymond and Muneri, 2001). Most often these were taken at breast height, but how representative these were of the entire tree was relatively unknown, with the exception of a few traits. Recent studies have aimed to clarify the variation in wood properties within a tree, and to devise the simplest and cheapest sampling strategies that will best reflect whole tree values for those wood properties. Within-tree variation in basic density, pulp yield and fibre properties has been explored extensively in several *Eucalyptus* species, including *E. globulus*, *E. regnans*, *E. nitens* and *E. fastigata* (Raymond et al., 1998; Schimleck and Michell, 1998; Kibblewhite and Riddell, 2000, 2001; Muneri and Raymond, 2001; Raymond and Muneri, 2001; Raymond et al., 2001b; Kube and Raymond, 2002; Kibblewhite et al., 2004). Within-tree variation in the chemical wood properties including lignin, extractives and cellulose has been investigated to a lesser extent and only in *E. globulus* and *E. nitens* (Pereira and Sardinha, 1984; Ona et al., 1997; Yokoi et al., 1999; Kibblewhite and Riddell, 2000; Kube and Raymond, 2002).

Significant within-tree variation in basic density, pulp yield and fibre properties has been found in *Eucalyptus*. Basic density generally decreases from the felling cut to the first sampling height (usually 10% of tree height), and then increases linearly until 70% of tree height in *E. globulus*, *E. nitens* and *E. fastigata* (Raymond and MacDonald, 1998; Evans et al., 2000; Raymond and Muneri, 2001; Kube and Raymond, 2002; Kibblewhite et al., 2004). The rate of increase in basic density

differs with species and was greatest for *E. nitens* (Raymond and MacDonald, 1998; Kibblewhite et al., 2004). For *E. regnans* no initial decline in basic density was observed, but rather a linear increase between 5 and 90% of tree height (Raymond et al., 1998). Density has also been shown to vary radially, decreasing from the pith and then increasing toward the bark in *E. nitens* (Greaves et al., 1997; Evans et al., 2000).

Reports of within-tree variation in pulp yield are less consistent. Parabolic patterns of change, where pulp yield increases from the base of the tree, reaches a plateau and then declines, have been found for *E. nitens* (Schimleck and Michell, 1998). Raymond et al. (2001b) found little change in pulp yield with height in *E. globulus*, whereas in *E. nitens* pulp yield remained constant up to 50% of tree height and then decreased. Pulp yield has also been found to vary radially, increasing from pith to bark (Schimleck and Michell, 1998). Within-tree patterns of variation for fibre length also differ between studies. Fibre length has been found to increase from the base of the tree, reach a plateau and then decrease in *E. nitens*, *E. globulus* and *E. regnans* (Raymond et al., 1998; Muneri and Raymond, 2001), whereas other studies have found a slight increase from the base followed by a sharp decline in *E. nitens* and *E. fastigata* (Kibblewhite and Riddell, 2000, 2001). Fibre coarseness decreased with height in *E. nitens*, *E. globulus* and *E. fastigata* (Muneri and Raymond, 2001; Kibblewhite et al., 2004), although in *E. regnans* it has been found to drop initially, stabilise between 15 and 70% of tree height and then decline at the top of the tree (Raymond et al., 1998). Microfibril angle has been found to follow a hyperbolic pattern of change in *E. nitens* and *E. fastigata*, decreasing from the base of the tree, reaching a plateau and then increasing with height, with the rate of change being much greater for *E. nitens* (Evans et al., 2000; Kibblewhite et al., 2004). Microfibril angle also varies radially, decreasing from pith to bark in *E. nitens* (Evans et al., 2000).

The pattern of within-tree variation often differs between the traits, species and studies, and therefore each must be considered individually to develop an effective and representative sampling strategy. The development of sampling strategies must also address issues such as whether whole tree, disks or increment wood cores should be sampled, and at which height they should be taken from. The sampling strategies that have been suggested vary between traits, species and sites, but in

general measuring a wood property from increment wood cores has been reported to give a reliable prediction of whole tree values (Raymond et al., 1998; Muneri and Raymond, 2001; Raymond and Muneri, 2001; Raymond et al., 2001b; Kube and Raymond, 2002). For *E. globulus* a consistent sampling strategy was developed for measuring basic density, pulp yield, fibre length and fibre coarseness by taking cores at a fixed height of 1.1 m (Muneri and Raymond, 2001; Raymond and Muneri, 2001; Raymond et al., 2001b). For *E. nitens*, sampling heights varied between traits and the quality of sites, from a fixed height of 0.7 m for basic density and fibre length to 0.9 m for fibre coarseness, basic density and pulp yield on good sites (Muneri and Raymond, 2001; Raymond and Muneri, 2001; Raymond et al., 2001b; Kube and Raymond, 2002). Optimal sampling height for pilodyn penetration was determined to be 1.3 m for *E. globulus* and 1.5 m for *E. nitens* (Raymond and MacDonald, 1998).

Within-tree variation in traits such as lignin, cellulose and extractives contents has been less well studied, and optimal sampling strategies have not been fully determined. These traits are usually measured from wood cores taken at breast height, but how representative these are of the whole tree is relatively unknown for *E. globulus*. Reports have generally been limited to fewer eucalypt species and to small sample sizes for some traits. The most comprehensive of these studies examined height variation in cellulose and extractives contents in 12 to 18 *E. nitens* trees, with significant differences found for both traits (Kube, 2005). Cellulose content showed a parabolic pattern of change with height, whereas extractives content was greatest at the base of the tree and decreased with height. Wood cores taken at a height of 0.9 m were found to give a good representation of whole tree values for both traits, although the relationship was not as strong for extractives content. Kibblewhite and Riddell (2000, 2001) examined variation in both lignin and cellulose contents with height in nine trees each of *E. nitens* and *E. fastigata*. Lignin content changed with height in a hyperbolic pattern, whereas cellulose content showed an opposite trend with a small increase from the base followed by a marked decline. In *E. globulus*, height variation in cellulose, lignin and extractives contents has been examined for ten trees (Pereira and Sardinha, 1984). Extractives content decreased with height, while lignin and cellulose contents remained relatively constant up to 70% of tree height. These studies only examined

longitudinal variation in the wood chemistry and radial variation was not explored. The only study looking at radial variation in the chemical wood traits examined two trees each of *E. globulus* and *E. camaldulensis* for variation in lignin content and the syringyl/guaiacyl (S/G) ratio of the lignin (Ona et al., 1997). In *E. globulus* lignin content had only radial variation, increasing from pith to bark, with no height variation observed. Patterns of variation in the S/G ratio were not consistent, with different patterns found for the two *E. globulus* trees. *E. camaldulensis* had almost no within-tree variation in lignin content, and the S/G ratio was generally higher in the pith than towards the bark, and higher at the top of the tree than the base. No study of *E. globulus* has yet incorporated both radial and longitudinal variation of the main chemical wood properties in the same set of trees and in a significant sample size.

The pith of trees that are under mechanical stress will often be located eccentrically inside the stem (Plomion et al., 2001). In eucalypts this is frequently associated with the presence of tension wood, which forms on the upper side of a leaning stem (Hillis, 1984). Tension wood is characterised by a different chemical composition to that of normal wood, with low lignin and extractives contents, and a high cellulose content. The impact of eccentric growth on the chemical composition of eucalypt wood and how this affects sampling has been relatively unexplored. Rodrigues et al. (2001) found no significant differences in the S/G ratio with eccentric growth in *E. globulus*, while significant differences were observed for lignin content in one out of three disks. This indicated that the presence of eccentric growth may need to be addressed during the development of sampling strategies for wood chemistry analyses in *E. globulus*.

Evaluation of within-tree variation in wood chemistry has been limited by the expensive and time consuming nature of wet chemistry measurements for lignin, cellulose and extractives contents. The development of near infrared reflectance (NIR) calibrations from which these traits can be predicted reliably, quickly and inexpensively in large numbers of samples, from both ground and solid wood, will allow these studies to be conducted in larger sample sizes (see Chapters 2 and 3; Raymond and Schimleck, 2002; Poke et al., 2004). This study explored the within-tree variation (both radially and longitudinally) in cellulose, lignin and extractives

contents in nine trees of *E. globulus* using NIR predictions. The impact of eccentric growth on radial variation in these traits was also evaluated. Knowledge of within-tree variation in the chemical wood traits would indicate if existing sampling strategies involving wood cores taken at breast height are effectively representing the whole tree. Correlations between these traits within-tree were determined.

Materials and Methods

Wood samples

The same nine 14-year old *E. globulus* trees from the West Ridgley base population trial (Gunns Ltd) described in Chapter 3, were also used in this study (Table 1). Bark-to-pith strips were taken at each 10% increment of total tree height from a constant direction, for NIR analysis and wood chemistry measurements. If eccentric growth was observed in the trees, bark-to-pith strips were also taken from the opposite side for comparison. NIR spectra were taken in increments along the transverse face of the strip (40 mm, followed by 20 mm increments) as detailed in Chapter 3. The extractives, lignin and cellulose contents were measured (Appita Standards, 1978, 1994; Wallis et al., 1997) for 40, 43 and 39 increment samples, respectively, for calibration development. These calibrations were further validated using an independent set of samples (see Chapter 3) and the extractives, lignin and cellulose contents predicted for the remaining increments. Acid-soluble lignin content measurements were excluded from the analysis because of the lack of accuracy of the predictions.

Statistical analysis

Analyses of variance were calculated using GenStat 6.1 (Lawes Agricultural Trust, 2002) with height, bark-to-pith position at each height, and radial side as fixed effects. Analyses of variance for height and bark-to-pith variation excluded eccentric samples taken from the opposing side of the tree. Probability values less than 0.05 were significant.

Table 1. *Eucalyptus globulus* samples from the West Ridgley base population trial used for within-tree analysis of lignin, cellulose and extractives contents. Lignin, extractives and cellulose measurements presented are averages from increment wood cores (detailed in Chapter 2; Poke et al., 2004; Apiolaza et al., in press).

Tree	Subrace	Locality	Family	Part of trial	Lignin content (%)	Extractives content (%)	Cellulose content (%)
1	Strzelecki Ranges	Bowden Road	319	North	26.14	3.74	43.4
2	Strzelecki Ranges	Bowden Road	322	South	27.92	4.07	-
3	Strzelecki Ranges	Bowden Road	322	South	28.43	3.43	-
4	Flinders Island	Central Flinders Island	426	North	29.91	6.04	40.3
5	Flinders Island	Central Flinders Island	426	North	28.92	5.90	42.7
6	Flinders Island	Central Flinders Island	427	North	29.12	8.98	40.4
7	Flinders Island	Central Flinders Island	427	North	29.59	10.28	39.7
8	North-eastern Tasmania	German Town	970	North	28.93	9.77	-
9	North-eastern Tasmania	German Town	970	North	28.79	3.96	-

Results

Trait statistics

The mean value, minimum and maximum for each trait is detailed in Table 2. Extractives content showed a large range of 13.94% (0.36% to 14.30%) and a mean of 4.45% indicating that very high extractives values were relatively rare. Klason lignin content had a range of 9.01% (15.69% to 24.70%) and a mean of 22.22% suggesting that most values were at the higher end of this scale. Total lignin content ranged from 21.39% to 30.18% (8.79%) with a mean of 28.20%. Cellulose content had a large range from 37.91% to 53.05 % (15.14%) and a mean of 42.90%.

Table 2. Statistics for extractives, Klason lignin, total lignin and cellulose contents within nine trees of *Eucalyptus globulus* based on near infrared predictions.

Trait	Mean (%)	Minimum (%)	Maximum (%)
Extractives content	4.45	0.36	14.30
Klason lignin content	22.22	15.69	24.70
Total lignin content	28.20	21.39	30.18
Cellulose content	42.90	37.91	53.05

Correlation between traits

Individual cellulose content measurements were highly, negatively correlated with both total lignin (R^2 = -0.70; Figure 1) and Klason lignin contents (R^2 = -0.63). Extractives content was poorly correlated with both cellulose (R^2 = -0.07) and total lignin (R^2 = 0.10) contents. Average bark-to-pith Klason lignin, total lignin and cellulose contents at each percentage tree height were poorly correlated with diameter at that percentage tree height (R^2 = 0.17, 0.01, and -0.001, respectively). Average bark-to-pith extractives content at each percentage tree height was highly, positively correlated with diameter at each percentage tree height (R^2 = 0.62; Figure 2).

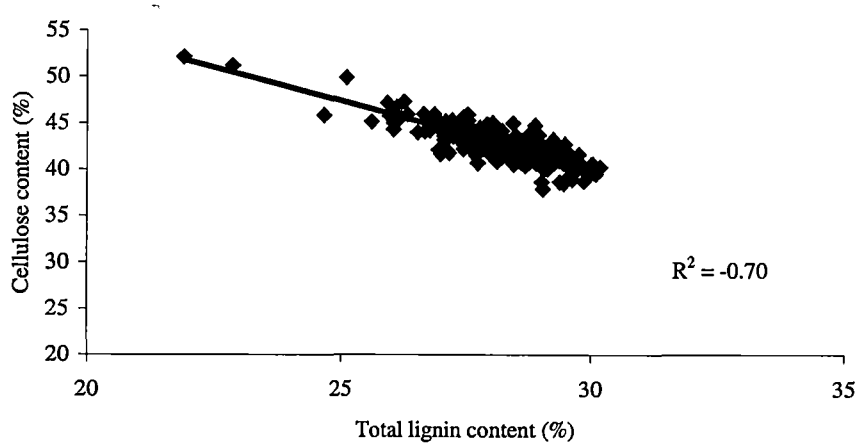

Figure 1. Correlation between total lignin content and cellulose content within nine trees of *Eucalyptus globulus*.

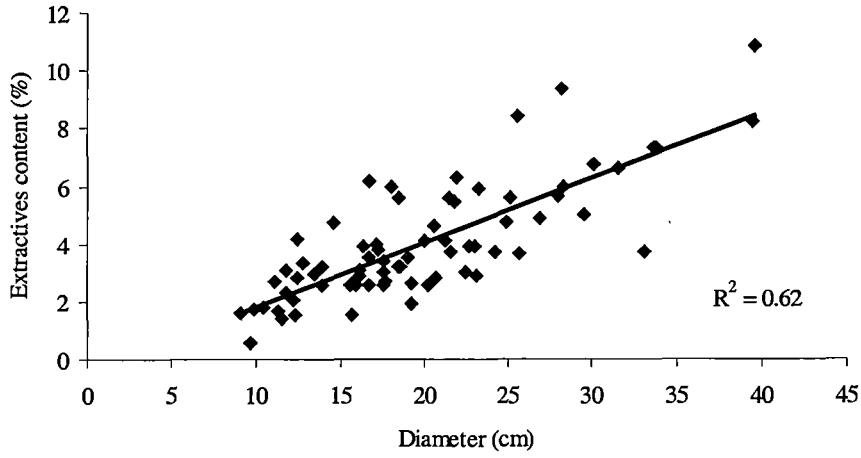

Figure 2. Correlation between extractives content and diameter within nine trees of *Eucalyptus globulus*.

Bark-to-pith variation in extractives, lignin and cellulose contents

Extractives content showed no significant bark-to-pith variation (Table 3; Figure 3). Cellulose content showed significant bark-to-pith variation at one height only (40%) (Table 3), although the trend was for cellulose content to decrease non-linearly from bark-to-pith at each height (Figure 4). No significant bark-to-pith variation was found for both Klason lignin and total lignin contents up to 10%, however, from 20% to 60% of tree height, both showed significant bark-to-pith variation (Table 3). No variation was found at 70% (Table 3). Both Klason lignin and total lignin contents generally increased non-linearly from bark-to-pith (Figures 5 and 6).

Height variation in extractives, lignin and cellulose contents

Height was found to be a significant source of variation for extractives content ($P < 0.001$) with the average content decreasing non-linearly with height in the tree (Figure 7). Klason lignin ($P < 0.06$), total lignin ($P < 0.97$) and cellulose ($P < 0.16$) contents showed no significant variation with height (Figure 8).

Table 3. Significance tests for bark-to-pith variation in extractives, Klason lignin, total lignin and cellulose contents within nine trees of *Eucalyptus globulus*. Tree height is the average height for all nine trees at which disk samples were taken. Numbers in the last four columns are F values. Significant probability values are denoted ***P < 0.001, **P < 0.01, *P < 0.05 and ns = non-significant.

Percentage tree height (%)	Tree height (m)	df	Extractives content	Cellulose content	Klason lignin content	Total lignin content
0	0	8	1.03 ns	1.21 ns	1.72 ns	1.00 ns
10	2.49	4	1.48 ns	0.50 ns	1.78 ns	1.47 ns
20	4.98	6	1.57 ns	2.38 ns	2.60 *	2.51 *
30	7.47	4	2.45 ns	2.15 ns	9.13 ***	10.02 ***
40	9.96	4	1.01 ns	5.66 **	10.76 ***	14.83 ***
50	12.45	3	2.02 ns	1.45 ns	9.50 ***	12.30 ***
60	14.94	2	2.67 ns	0.85 ns	3.77 *	5.89 **
70	17.43	2	0.13 ns	0.16 ns	2.43 ns	3.61 ns

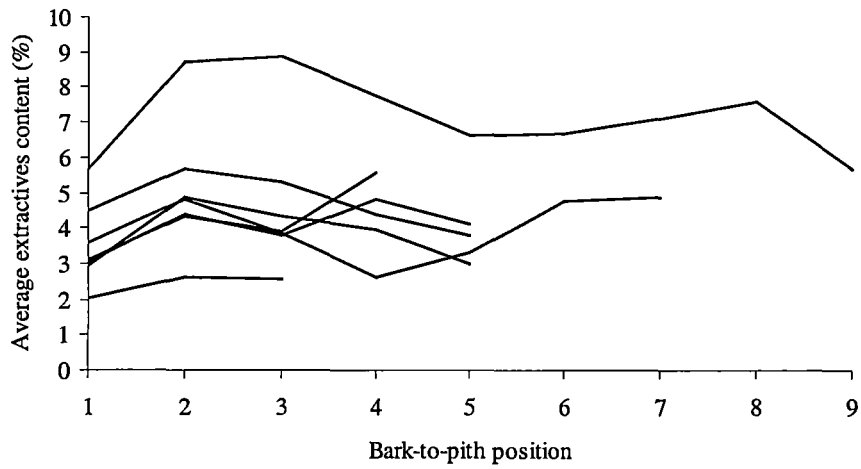

Figure 3. Bark-to-pith variation in extractives content from 0 to 60% of tree height within nine trees of *E. globulus*. The top line represents 10% of tree height with the percentage tree height increasing in sequence down the graph.

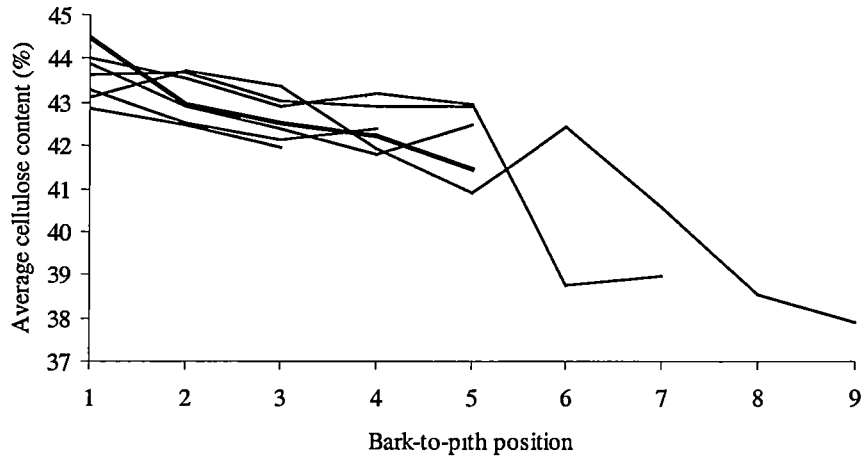

Figure 4. Bark-to-pith variation in cellulose content from 0 to 60% of tree height within nine trees of *E. globulus*. The bold line represents bark-to-pith variation at 40% tree height, the only height at which radial variation was significant.

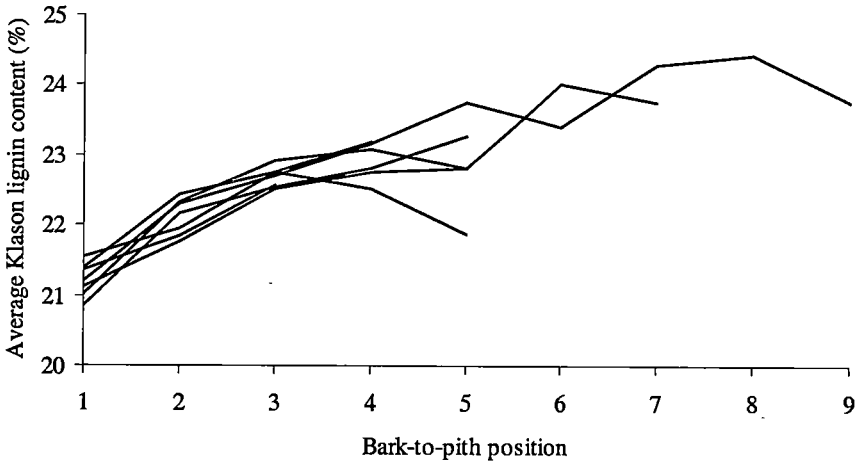

Figure 5. Bark-to-pith variation in Klason lignin content from 0 to 60% of tree height within nine trees of *E. globulus*.

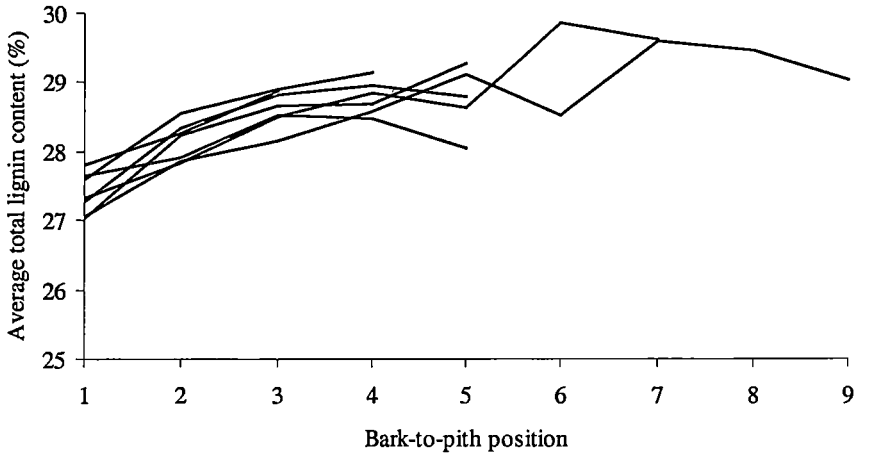

Figure 6. Bark-to-pith variation in total lignin content from 0 to 60% of tree height within nine trees of *E. globulus*.

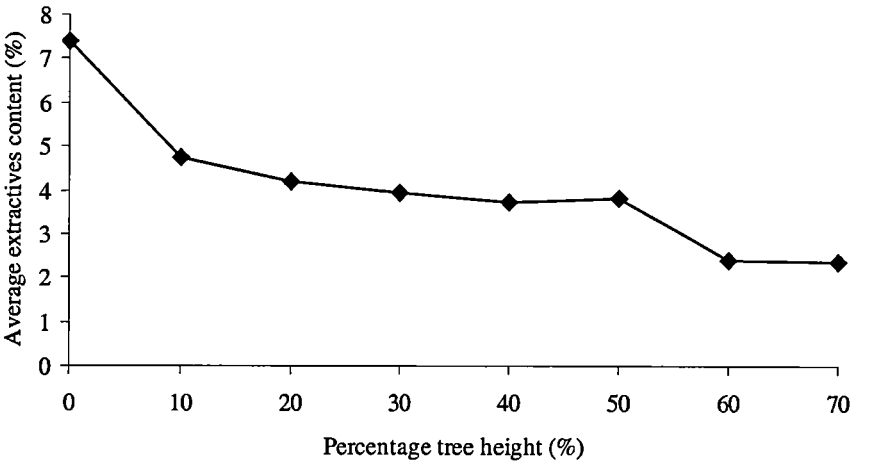

Figure 7. Change in extractives content with height within nine *Eucalyptus globulus* trees.

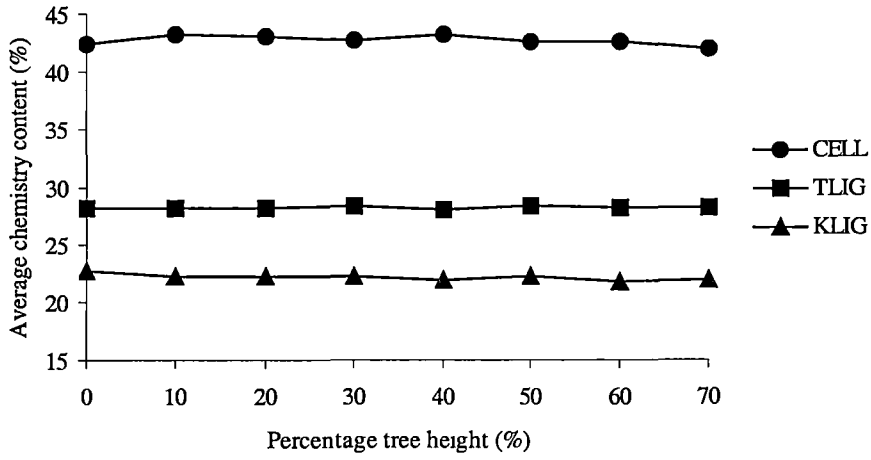

Figure 8. Change in cellulose (CELL), total lignin (TLIG) and Klason lignin (KLIG) contents with height within nine *Eucalyptus globulus* trees.

Variation in extractives, lignin and cellulose contents with eccentric growth

Eccentric growth was observed in seven disk samples taken from four of the nine trees. Extractives content showed no variation between radial sides with eccentric growth ($P < 0.97$). Klason lignin ($P < 0.002$), total lignin ($P < 0.002$) and cellulose ($P < 0.008$) contents varied significantly between radial sides.

Discussion

Within-tree variation was identified for lignin, cellulose and extractives contents in nine trees of *E. globulus*, including both longitudinal and radial variation. Extractives content had no significant bark-to-pith variation but significantly decreased with height in the tree. Cellulose content showed significant radial variation at 40% of tree height (approximately 10 m) only, whereas Klason lignin and total lignin content showed significant radial variation between 20% and 60% of tree height (5 m to 15 m). The general bark-to-pith trend was for cellulose content to decrease and lignin content to increase. Both lignin and cellulose contents showed no significant longitudinal variation within-tree. This is the first report of within-tree variation in these wood chemistry components of *E. globulus* that includes the evaluation of both radial and longitudinal variation in the same set of trees and for a sample size larger than two trees.

Extractives content decreased with height in the nine *E. globulus* trees, supporting the findings of Pereira and Sardinha (1984) for ten *E. globulus* trees and of Kube (2005) for *E. nitens*. In agreement with this a strong, positive correlation was found between extractives content and diameter (R^2 = 0.62). Longitudinal variation in extractives content is similar in both *E. globulus* and *E. nitens*, and shows a comparable pattern of variation to fibre coarseness, which has been reported for *E. globulus, E. nitens* and *E. fastigata* (Muneri and Raymond, 2001; Kibblewhite et al., 2004). Significant height variation in extractives content suggests that increment wood cores taken at breast height are unlikely to reflect whole tree measurements. Kube (2005) found that cores taken at a recommended height of 0.9 m in *E. nitens* only explained 56% of the whole tree variation in extractives content. For *E. globulus*, cores taken between 20 and 50% of tree height (approximately 5 – 12 m), where extractives content is more stable, are likely to give better predictions of whole tree values. Extractives need to be removed during kraft pulping and are therefore undesirable. Cores taken lower in the trunk where extractives content is greatest will give a maximum value rather than an average, which will be useful when screening for trees with low extractives content.

No significant bark-to-pith variation was found for extractives content at any height, which is the first evaluation of radial variation for this trait in *E. globulus*. Extractives content has been reported to be low in the sapwood of *Eucalyptus*, and in the heartwood, decrease from the outer heartwood to the pith (see Hillis, 1984). In this study the sapwood had consistently low levels of extractives, but the increase from the pith to the outer heartwood was only observed for three non-consecutive heights, while the other four heights showed an initial decrease followed by an increase (Figure 3). The absence of significant bark-to-pith variation means that the outer wood or partial cores could be sampled, which would be less destructive to the tree. Extractives content was poorly correlated with cellulose and lignin contents indicating these traits would have a different pattern of within-tree variation. It also suggested that the genetic control of the distribution of these chemical components is quite different. Ona et al. (1998) found highly significant correlations between these traits within *E. globulus* but with a much smaller sample size, which means there may be considerable environmental influences.

Two previous studies using ten and two trees, respectively, found no variation in lignin content with tree height in *E. globulus* (Pereira and Sardinha, 1984; Ona et al., 1997). The current study is in agreement with those findings showing no variation with height for both Klason lignin and total lignin contents. Lignin content has been reported to have a hyperbolic pattern of longitudinal variation in *E. nitens* and *E. fastigata*, initially decreasing from the base and then increasing with height (Kibblewhite and Riddell, 2000, 2001). These dissimilarities between the eucalypt species demonstrate the inherent differences in wood chemistry and its genetic control between species. Taking wood cores at breast height seems to be a suitable sampling strategy for lignin content measurements in *E. globulus*, and may be reflective of whole tree values, unlike *E. nitens* or *E. fastigata*. Significant radial variation in lignin content was found in the nine *E. globulus* trees between 20 and 60% of tree height (5 to 15 m), with an increase from the bark to the pith. This trend is the opposite to that found by Ona et al. (1997) in two *E. globulus* trees, which may be due to differences in methods used or in the precision of the methods. This suggests that larger sample sizes are needed for studying within-tree variation in wood chemistry. The radial variation in lignin content found here is consistent with an opposite trend found for pulp yield in *E. nitens* (Schimleck and Michell, 1998), a trait that is highly, negatively correlated with lignin content (Chapter 5; Wallis et al., 1996). The same pattern of radial variation has been found for microfibril angle (Evans et al., 2000) as that found here for lignin content. These traits have been found to be highly, positively correlated (see Chapter 5), indicating there may be some interaction between them. Radial variation in lignin content only occurred higher up in the tree, suggesting that partial cores may be sampled at breast height for trait analysis, which may reduce the sampling time and detriment to the tree.

Within-tree measurements for lignin and cellulose content were highly, negatively correlated ($R^2 = -0.70$), which supported the highly significant negative correlations obtained by Ona et al. (1998) within *E. globulus*. This suggested that patterns of within-tree variation in cellulose content would be complementary to that found for lignin content. Similar to lignin content, cellulose content showed no significant variation with tree height in *E. globulus*. This was consistent with the results of Pereira and Sardinha (1984), who found no longitudinal variation in cellulose

content in *E. globulus*, and of Raymond et al. (2001b) who found no variation in pulp yield with height in *E. globulus*, a trait highly correlated with cellulose content (Wallis et al., 1996; Raymond and Schimleck, 2002; Apiolaza et al., in press). Cellulose content has been found to vary with tree height in other eucalypt species including *E. nitens* and *E. fastigata* (Kibblewhite and Riddell, 2000, 2001; Kube, 2005). For these species a parabolic pattern of variation was observed, with cellulose content increasing from the base of the tree and then decreasing with height, an opposite trend to that found for lignin content in these species (Kibblewhite and Riddell, 2000, 2001). Once again the inherent differences in wood chemistry and its genetic control between the eucalypt species are highlighted. Wood cores taken at breast height are likely to reflect whole tree values for cellulose content in *E. globulus*, similar to pulp yield (Raymond et al., 2001b), due to no longitudinal variation in this trait. Despite cellulose content generally decreasing from bark-to-pith (opposite to lignin content) it only showed significant radial variation at 40% of total tree height. Radial patterns of variation in cellulose content have not been examined before in *Eucalyptus*, although, pulp yield has been found to decrease bark-to-pith in *E. nitens* (Schimleck and Michell, 1998), suggesting cellulose content may be consistent with this. This pattern of variation is opposite to what has been found for microfibril angle in *E. nitens* (Evans et al., 2000), suggesting these traits may be highly correlated. Similar to lignin content it is likely that the outer wood sampled at low heights (less than 2 m) will be sufficient for cellulose content analyses.

The effect of eccentric growth on wood chemistry was observed by comparing extractives, lignin and cellulose contents between the short and long radial sides of a bark-to-bark strip. Eccentric growth was identified in seven disk samples, for which there were significant differences between sides for lignin and cellulose contents, but not for extractives contents. The impact of eccentric growth on cellulose and extractives contents has not been examined before for *Eucalyptus*. Rodrigues et al. (2001) reported some evidence of significant variation in lignin content with eccentric growth in *E. globulus*, but only in one out of three disk samples. The differences in wood chemistry between the radial sides, suggests that tension wood may be present, which has low lignin and high cellulose contents (Plomion et al., 2001; Rodrigues et al., 2001). Therefore eccentric growth needs to

be considered when sampling for lignin and cellulose analyses, but will be less problematic for extractives content.

Within-tree variation in wood chemical composition was found for nine trees of *E. globulus*. Extractives content showed no radial variation, but decreased with tree height, suggesting wood cores sampled at breast height may better reflect maximum values rather than whole tree values. Lignin and cellulose contents did not vary with height indicating wood cores taken at breast height would be optimal. Lignin and cellulose contents varied bark-to-pith at heights greater than 5 m, with the general trend for an increase and decrease, respectively. Despite this, sampling the outer wood could be sufficient when measuring all of the above traits at breast height. Eccentric growth was observed to be a factor in within-tree variation in lignin and cellulose contents, but not for extractives content, suggesting it needs to be considered when sampling for the former two traits. Present sampling strategies for lignin and cellulose contents in *E. globulus* were found to be adequate and likely to give an accurate representation of whole tree values. For extractives content, revision of the sampling height may be required if values representative of the whole tree are desired.

CHAPTER 5

Genetic parameters for lignin, extractives and decay in *Eucalyptus globulus*

This chapter has been accepted for publication as: Poke F.S., Potts B.M., Vaillancourt R.E., Raymond C.A. (in press) Genetic parameters for lignin, extractives and decay in *Eucalyptus globulus*. Annals of Forest Science.

Introduction

Eucalyptus globulus is grown for pulp production in temperate Australia and other parts of the world, including South America, southern Europe, Africa and Asia (Eldridge et al., 1993). Considerable genetic variation has been identified between the subraces of *E. globulus* for a wide range of traits, including growth and both physical and chemical wood properties (Dutkowski and Potts, 1999; Miranda and Pereira, 2001, 2002; Apiolaza et al., in press). Some of this variation has been exploited in breeding programs for the selection of superior trees. When selecting trees for pulp production, only a few traits are currently examined in Australia, with the focus on increases in volume per hectare, basic density and pulp yield (Greaves and Borralho, 1996; Greaves et al., 1997; Apiolaza et al., in press). Although selection for these traits gives an increase in the pulp yield per hectare, many other physical and chemical wood properties are important to kraft pulping, and variations in these can be conducive to minimising the costs or environmental impacts of the process.

Kraft pulping generally involves the removal of most of the extractives, approximately 80% of the lignin and approximately 50% of the hemicellulose from the cellulose fibres using alkali (Smook, 1992). For the production of high quality paper, the pulp is further bleached to remove the residual lignin, which is responsible for turning the paper yellow through oxidation and light absorption

(Smook, 1992). The lignin and extractives contents of wood, are traits that are fast being recognised as having importance in minimising the costs and environmental impacts of kraft pulping. As lignin and extractives are the primary waste products of the pulping process, lower levels in the wood will result in faster delignification and/or a reduction in the use of chemicals and energy. This will help minimise the production of pollutants from the pulping process.

Studies into the genetic variation and heritability of lignin and extractives have been limited in *Eucalyptus* until recently. This was mainly due to the expensive and time-consuming nature of the chemical assay used to measure these traits (Appita Standards, 1978, 1994). More recently, simple and cost-effective techniques have been developed for predicting these traits, involving near infrared reflectance (NIR) analysis of ground wood cores (Baillères et al., 2002; Poke et al., 2004; Chapter 2). This has been found to be an effective technique for reliably predicting these traits in large numbers of samples. A small number of studies have suggested that there is genetic variation in lignin and extractives contents in *E. globulus*, although these involved only three or four provenances and a small number of individuals (Washusen et al., 2001; Miranda and Pereira, 2002). Studies investigating the correlations between the chemical wood traits and other wood properties in *E. globulus*, have also been limited by small sample sizes and to small numbers of traits, and have involved phenotypic correlations only (Wallis et al., 1996; Miranda and Pereira, 2001). Genetic correlations have not been reported for these traits in *E. globulus*. To fully explore the scope of variation within the chemical traits and their genetic control, large numbers of individuals, families and provenances are required which encompass the range of *E. globulus*.

The susceptibility of trees to wood decay is important as it may impact on pulp yield in two ways: firstly, the plant defence response may lead to an increase in the amount of lignin and extractives present in the wood, which will reduce the pulp yield; secondly, decay leads to degradation of the wood causing a breakdown of the cellular structure (Pearce, 1996). This decay can be caused by pathogen infection of exposed, vulnerable tissue following wounding, or through attack of the heartwood (heart rot), which is incapable of an active response due to the lack of living cells (Pearce, 1996). The impact of decay will also depend upon whether the decay

organisms are feeding on cellulose or lignin. Two types of fungi generally are responsible for decay, brown rot fungi which degrade cellulose and white rot fungi which degrade lignin. Decay can be observed in "pockets" that are compartmentalised by a reaction zone (purple/pink coloured boundary between healthy and decayed wood) and discolouration of the surrounding wood, or as rotting of the heartwood (Pearce, 1996; Wardlaw et al., 2003). Fungal attack has been found to be associated with increases in lignin due to its resistance to degradation by pathogens, in *E. gunnii* (Hawkins and Boudet, 2003), and increases in extractives which contain antimicrobial compounds, in *E. nitens* (Barry et al., 2001). In *E. globulus* genetic variation in decay susceptibility and its relationship to other wood properties has not been examined.

Improving the chemical wood properties of tree species through breeding requires natural variation to be present for selection. It is also necessary to understand how the traits to be improved are related to one another and to other traits that are not currently part of the breeding program, so that when an increase in one trait is selected for, the potential effects on other traits may be predicted. A study conducted by Apiolaza et al. (in press) examined the variation in growth and wood traits as well as their correlations using 188 individuals of *E. globulus* from 35 families and eight subraces, which are currently part of the single breeding population of the Australian national breeding program. The traits examined included diameter at breast height over bark (DBH), basic density (BD), mean fibre length (FL), mean microfibril angle (MFA) which is the average angle of the cellulose microfibrillar helix relative to the longitudinal fibre axis (Evans et al., 1999), predicted pulp yield (PY) and cellulose content (CELL). The current study aimed to build on that of Apiolaza et al. (in press) with a particular focus on the chemical wood properties. Using the same open-pollinated progenies grown in a field trial, we examined the variation in and the heritability of lignin content (LIG), extractives content (EXTR) and extent of decay (DEC) between and within nine localities of *E. globulus* originating from around Tasmania and south-east Victoria. Phenotypic and genetic correlations were also determined amongst these traits and with the growth and wood traits of Apiolaza et al. (in press). The relationships between these chemical wood traits and with the physical wood traits, wood decay

and growth has not been examined before in *E. globulus* and will provide an indication of how multiple traits are affected during the selection of superior trees.

Materials and Methods

Plant material

Wood samples of *Eucalyptus globulus* were collected from a base population field trial located at West Ridgley, Tasmania (Gunns Ltd). This trial was established in 1989 based on the CSIRO Australian Tree Seed Centre collection and is comprised of open-pollinated families (Gardiner and Crawford, 1987, 1988). The trial was an incomplete block design with 451 families in five replicates, each with 17 incomplete blocks, and two-tree row plots (Apiolaza et al., in press). A total of 177 trees from 37 families (Table 1) were sampled to cover the same range of eight subraces sampled by Apiolaza et al. (in press), with one tree or occasionally two trees per plot sampled. Due to the fact that only a subset of families was sampled the trial was treated as a randomised complete block design for analysis. The locality denoted North-east Tasmania comprises two localities, Royal George and German Town, which were merged because of small sample sizes and their close proximity. Two bark-to-bark wood cores were taken from each tree approximately 10 cm above the previous core sites taken by Apiolaza et al. (in press) two years before, according to the method described by Raymond et al. (2001b).

Wood and growth measurements

Measurements for BD, MFA, FL, PY and CELL already existed for these trees at age 11 years (Apiolaza et al., in press). Additional measurements were taken for BD and DBH, and measurements were obtained for DEC, LIG and EXTR all at age 13 years.

DEC was recorded for each core as the percentage of the core with heart rot, pocket decay and/or discolouration and results were averaged for the tree. For statistical analysis the different types of decay data were grouped, with 0 indicating no decay followed by 10% intervals thereafter, and class midpoints were used for analysis.

Due to the presence of decay in the pith for many of the cores, partial cores (the outer quarters of each core free of decay) were used for further BD, LIG and EXTR measurements.

Table 1. Breakdown of subraces (as classified by Dutkowski and Potts [1999]), localities and families of *E. globulus* used in this study from the base population trial at West Ridgley, Tasmania.

Subrace	Locality	Number of families	Number of individuals
Flinders Island, Tasmania	Central Flinders Island	4	18
King Island, Tasmania	South King Island	5	27
North-eastern Tasmania	North-east Tasmania	4	16
South-eastern Tasmania	Moogara	4	14
South-eastern Tasmania	North Maria Island	3	14
Southern Tasmania	South Geeveston	4	18
Strzelecki Foothills, Victoria	Madalya Road	4	20
Strzelecki Ranges, Victoria	Bowden Road	4	22
Western Otways, Victoria	Cannan Spur	5	28
Total	9	37	177

BD was determined for one core from each tree using the water displacement method (TAPPI, 1989), by submerging the partial cores in cold water for approximately two days, removal of remaining bark and excess water followed by volume (V) measurements. The mass (M) of each core was taken after drying at 105°C for approximately two days. BD was calculated using the following formula:

$$BD_{(kg/m^3)} = \frac{M}{V} \times 1000$$

The remaining partial cores were used to develop the NIR calibrations reported in Chapter 2 (Poke et al., 2004) for total lignin (TLIG), acid-soluble lignin (ASLIG) and Klason lignin (KLIG) contents (TLIG = ASLIG + KLIG) plus EXTR. These calibrations were based on chemical measurements for 54 to 61 samples and had good correlation coefficients (0.62-0.93), and predicted and laboratory values for the validation set of samples were highly correlated (0.83-0.99) (Poke et al., 2004;

Chapter 2). The calibrations were used to predict these traits for the remainder of the individuals in the data set.

Statistical analysis

Variance components for BD, DBH, ASLIG, KLIG, TLIG, EXTR and DEC were estimated using the MIXED procedure in SAS (Version 9.1, SAS Institute Inc.), with locality fitted as a fixed effect, and family within localities, replicate and residual within localities as random effects. Locality least square means and the differences between them were also calculated using the MIXED procedure in SAS, with a Tukey-Kramer adjustment applied for multiple comparisons.

The individual narrow-sense (h^2_{op}) and family means (H^2_{fm}) heritabilities of BD, DBH, ASLIG, KLIG, TLIG, EXTR and DEC were estimated using ASREML (Gilmour et al., 1999), with the fixed locality term removed from the model in the latter case. h^2_{op} refers to the narrow-sense heritability within localities, which is used operationally to predict genetic gains from within locality selection. H^2_{fm} is the family means heritability which indicates the gain that would be made from selecting the best families across all localities for deployment. h^2_{op} and H^2_{fm} were estimated as:

$$h^2_{op} = \frac{\sigma^2_{add(loc)}}{\left(\sigma^2_{add(loc)} + \sigma^2_e\right)}$$

$$H^2_{fm} = \frac{\sigma^2_f}{\sigma^2_f + \sigma^2_e / k}$$ (Johnson and Burdon, 1990)

where: $\sigma^2_{add(loc)}$ = additive genetic variation within locality variance component estimated assuming a coefficient of relatedness within open-pollinated families of 0.4, after first adjusting the additive relationship matrix for a 30% selfing rate (Dutkowski et al., 2001)

σ^2_f = family variance component calculated across localities

σ_e^2 = residual variance component

k = harmonic mean number of trees per family

Trait correlations were determined amongst the age 13 measurements of BD, DBH, ASLIG, KLIG, TLIG, EXTR and DEC and with the traits of Apiolaza et al. (in press). Phenotypic correlations (Pearsons correlation matrix) amongst individuals were determined in SAS using the CORR procedure. Additive genetic correlations could not be estimated using ASREML (Gilmour et al., 1999) directly, as bivariate models failed to converge due to the small sample size. However, as an approximation of the genetic correlations, Pearsons correlation matrices were obtained between the family means adjusted for locality differences and also between the nine locality means, using the CORR procedure in SAS.

Results

Trait statistics and variances

The number of individuals measured for each trait and the statistics for each trait are detailed in Table 2 and include the subset of measurements from Apiolaza et al. (in press). 69% of samples were found to have decay symptoms. Variation in the traits measured in this study, between replicates, localities and family within localities, are detailed in Table 3. No significant variation was detected at any level for BD, TLIG and EXTR (Table 3). Locality was a significant source of variation for DEC, DBH, KLIG and ASLIG (Table 3). Of these four traits, only two had significant differences between the locality least square means following Tukey-Kramer adjustment (Table 4). For DEC, South King Island had significantly more decay than five other localities including Bowden Road ($P < 0.001$), Madalya Road ($P < 0.002$), Central Flinders Island ($P < 0.003$), North-east Tasmania ($P < 0.004$) and Cannan Spur ($P < 0.023$). For ASLIG, South Geeveston and South King Island were significantly different to each other ($P < 0.01$). Significant variation between families within locality was detected for ASLIG only.

Table 2. Statistics for growth and wood measurements of individual trees for the *E. globulus* base population trial at West Ridgley, Tasmania.

Trait (Abbreviation)	Unit	n	Mean	Standard deviation	Minimum	Maximum
Mean fibre length at age 11 (FL)	mm	141	0.77	0.06	0.59	0.95
Mean microfibril angle at age 11 (MFA)	°	149	16.9	2.9	11.7	27.5
Predicted pulp yield at age 11 (PY)	%	157	51.8	1.6	42.5	57.0
Cellulose content at age 11 (CELL)	%	157	42.4	1.5	37.8	46.6
Basic density at age 11 (BD)	kg/m^3	161	494.5	40.5	395.8	589.4
Basic density at age 13 (BD)	kg/m^3	133	522.9	44.9	412.1	667.7
Diameter at breast height at age 13 (DBH)	cm	177	24.1	5.4	13.4	37.5
Klason lignin content at age 13 (KLIG)	%	155	22.38	1.21	18.97	25.45
Acid-soluble lignin content at age 13 (ASLIG)	%	155	6.12	0.52	4.42	8.11
Total lignin content at age 13 (TLIG)	%	155	28.48	1.26	24.72	31.23
Extractives content at age 13 (EXTR)	%	155	6.00	1.84	2.12	12.73
Extent of decay at age 13 (DEC)	%	143	35.9	32.4	0	95.0

Table 3. Analyses of variance for growth and wood traits at age 13 years between replicates, localities, and families within localities, plus estimates of the heritability of within locality variation, and family means heritability, for these traits in the samples from the *E. globulus* base population trial at West Ridgley, Tasmania. Probability values are denoted ***P < 0.001, *P < 0.05 and ns = non-significant.

Trait	df	Basic density (BD)	Diameter at breast height (DBH)	Klason lignin content (KLIG)	Acid-soluble lignin content (ASLIG)	Total lignin content (TLIG)	Extractives content (EXTR)	Extent of decay (DEC)
Replicate Z value (Probability value)	4	0[a] (-)	0.57 (0.284) ns	1.26 (0.105) ns	0.92 (0 178) ns	1.24 (0.107) ns	0.64 (0.261) ns	0.40 (0.345) ns
Locality F value (Probability value)	8	1 83 (0.114) ns	2.73 (0.023) *	2 52 (0.033) *	2.52 (0.034) *	2.03 (0.079) ns	1.80 (0.120) ns	5.40 (0 0004) ***
Family [locality] Z value (Probability value)	28	0 89 (0 186) ns	0[a] (-)	0 688 (0.249) ns	1.90 (0.028) *	1.2 (0.115) ns	1.52 (0.064) ns	0[a] (-)
Narrow-sense heritability (standard error)		0.24 (0.26)	0[a]	0.13 (0.20)	0.51 (0.26)	0.29 (0.23)	0.35 (0.23)	0[a]
Family means heritability (standard error)		0.42 (0.19)	0.19 (0.19)	0.42 (0.16)	0.64 (0.10)	0.50 (0.14)	0.48 (0.14)	0 50 (0.14)

Z values are random terms and F values depict fixed terms
[a] variance component was at the boundary of the parameter space

Heritability estimates

Narrow-sense heritability estimates had large standard errors due to the lack of significant variation between families within localities for most traits, no doubt

reflecting the small sample size (Table 3). Moderately high heritability values were obtained for ASLIG (0.51 ± 0.26) and EXTR (0.35 ± 0.23), although ASLIG was the only trait where significant variation between families within localities was detected. Both BD (0.24 ± 0.26) and TLIG (0.29 ± 0.23) showed moderate heritabilities, with KLIG (0.13 ± 0.20) showing little heritability. Within locality variation in DEC and DBH was non-heritable. The heritabilities of family means integrated both within and between locality variation, and were somewhat higher than the narrow sense heritabilities due to the inclusion of locality effects in the differences between families. ASLIG (0.64 ± 0.10), EXTR (0.48 ± 0.14), TLIG (0.50 ± 0.14) and DEC (0.50 ± 0.14) showed high estimates. BD (0.42 ± 0.19) and KLIG (0.42 ± 0.16) had moderately high estimates, and DBH a moderate estimate (0.19 ± 0.19) (Table 3).

Table 4. Locality least square means and standard errors (in parenthesis) for growth and wood traits at age 13 years for samples from the *E. globulus* base population trial at West Ridgley, Tasmania.

Locality	Basic density (BD) (kg/m^3)	Diameter at breast height (DBH) (cm)	Klason lignin content (KLIG) (%)	Acid-soluble lignin content (ASLIG) (%)	Total lignin content (TLIG) (%)	Extractives content (EXTR) (%)	Extent of decay (DEC) (%)
Central Flinders Island	520 (12) a	26.0 (1.2) a	22.6 (0.4) a	6.0 (0.2) ab	28.6 (0.4) a	6.4 (0.6) a	23.4 (7.2) a
South King Island	486 (13) a	25.3 (1.0) a	21.9 (0.4) a	5.7 (0.1) a	27.7 (0.4) a	5 4 (0.5) a	65.6 (6.5) b
North-east Tasmania	512 (14) a	20.9 (1.3) a	22.7 (0.4) a	6.1 (0.2) ab	28.8 (0.4) a	7.1 (0.6) a	22.1 (7.9) a
Moogara	511 (14) a	22.2 (1.4) a	22.6 (0.4) a	6.3 (0 2) ab	28.9 (0.4) a	6.1 (0.6) a	48.8 (8.9) ab
North Maria Island	537 (16) a	23 2 (1.4) a	22.5 (0.4) a	6.0 (0.2) ab	28.5 (0.5) a	6.0 (0.7) a	40.2 (8.5) ab
South Geeveston	532 (14) a	25.6 (1.2) a	21.8 (0.4) a	6.6 (0.2) b	28.2 (0.4) a	5.0 (0.6) a	48 6 (8.5) ab
Madalya Road	535 (12) a	22.9 (1.2) a	22.8 (0 4) a	6.0 (0 2) ab	28.8 (0.4) a	6.7 (0.5) a	21.3 (7.4) a
Bowden Road	542 (12) a	22 7 (1.1) a	22.8 (0.4) a	6.2 (0.1) ab	28.9 (0 4) a	6.2 (0.5) a	18.6 (7.4) a
Cannan Spur	527 (11) a	26.3 (1.0) a	21.8 (0 3) a	6.2 (0.1) ab	27.9 (0.4) a	5.3 (0.5) a	33.7 (6.1) a

Localities with common letters for the same trait are not significantly different at P < 0.05 following Tukey-Kramer adjustment for multiple comparisons

Trait correlations

Strong correlations were identified between the wood and growth traits at locality, family and individual (phenotypic) levels (Table 5). Correlations between locality means and family means adjusted for locality differences represented genetic based correlations. As expected, a strong, positive relationship was identified between TLIG and its components (ASLIG and KLIG) at most levels, although KLIG and ASLIG were not significantly correlated. EXTR was strongly correlated with lignin content for individuals, but the correlations were positive with KLIG and TLIG, and negative with ASLIG. Genetic correlations were observed between EXTR and both KLIG (families and localities) and TLIG (localities). KLIG, TLIG and EXTR all had significant, negative, phenotypic and genetic correlations with both CELL and PY, although these were sometimes not significant at the locality level. Lignin content showed significant negative correlations with BD at the family level supported at both ages 11 and 13 years. TLIG was also positively correlated with MFA at the individual and family level, with ASLIG and KLIG correlated with MFA at the individual level only. TLIG and DBH showed a weak negative correlation at the locality level only.

DEC was highly negatively correlated with BD only at age 11 years for individuals and localities, but not for families. Significant genetic variation has been reported for BD at age 11 at the subrace level (Apiolaza et al., in press). When the South King Island locality (particularly susceptible to decay) was removed from the analysis, DEC and BD (age 11) were no longer correlated at the locality level, but a significant correlation still remained at the individual level ($r = -0.230$, $P < 0.017$). DEC had weak, positive phenotypic correlations with PY and CELL, but there were no significant genetic relationships. DEC showed negative relationships with KLIG, TLIG and EXTR for localities, which were still significant for KLIG ($r = -0.714$, $P < 0.047$) and EXTR ($r = -0.717$, $P < 0.045$) when South King Island was removed from the analysis. DEC had a positive correlation with FL at the family level only.

Table 5. Correlations (Pearsons correlation matrix) amongst growth and wood traits for the *E. globulus* base population trial at West Ridgley, Tasmania. L = correlations amongst the nine locality means (df = 7), F = correlations amongst family means (adjusted for locality differences; df = 25 to 27) and I = phenotypic correlations amongst individuals (df = 106 to 155). Significant probability values are denoted ***P < 0.001, **P < 0.01, *P < 0.05.

Trait	Type	Klason lignin content (KLIG) (age 13)	Acid-soluble lignin content (ASLIG) (age 13)	Total lignin content (TLIG) (age 13)	Extractives content (EXTR) (age 13)	Extent of decay (DEC) (age 13)
Mean fibre length	L	-0.576	0.558	-0 308	-0.553	0 019
(FL)	F	-0 293	-0.329	-0.353	-0.179	0.386 *
(age 11)	I	-0.156	0.082	-0.124	-0.185 *	0.110
Mean microfibril	L	0.109	0.549	0.397	0.140	0.049
angle (MFA)	F	0.372	0.281	0 408 *	0.312	-0.005
(age 11)	I	0.299 ***	0.198 *	0.361 ***	0.188 *	-0 024
Predicted pulp	L	-0.707 *	0.314	-0.552	-0.779 *	0.530
yield (PY)	F	-0.690 ***	-0.175	-0.639 ***	-0.368	0.027
(age 11)	I	-0 426 ***	-0.023	-0.421 ***	-0.379 ***	0.221 *
Cellulose content	L	-0.640	0.204	-0.550	-0.744 *	0.183
(CELL)	F	-0.592 **	-0.199	-0.553 **	-0.451 *	0.006
(age 11)	I	-0.401 ***	-0.028	-0.394 ***	-0.399 ***	0.183 *
Basic density	L	0.367	0.284	0.495	0.200	-0 704 *
(BD)	F	-0.417 *	-0.500 **	-0.540 **	-0.195	-0.091
(age 11)	I	-0.007	-0.079	-0.055	0.057	-0.339 ***
Basic density	L	0 235	0.153	0.307	0.152	-0.460
(BD)	F	-0.415 *	-0.419 *	-0.510 **	0.227	0.264
(age 13)	I	-0.183 *	-0.074	-0 225 **	0.189 *	-0.032
Diameter at breast	L	-0 637	-0.075	-0.674 *	-0.618	0.286
height (DBH)	F	0.279	0.107	0.283	0.101	-0.055
(age 13)	I	0.069	0.047	0.103	-0.003	0 087
Klason lignin	L		-0.275	0.874 **	0.907 ***	-0.706 *
content (KLIG)	F		0.226	0.939 ***	0.442 *	-0 251
(age 13)	I		-0.141	0.930 ***	0.533 ***	-0 044
Acid-soluble lignin content (ASLIG) (age 13)	L			0.226	-0.321	-0.115
	F			0.543 **	-0.162	-0.295
	I			0.226 **	-0.312 ***	-0.089
Total lignin	L				0.752 *	-0.763 *
content (TLIG)	F				0.306	-0.322
(age 13)	I				0.396 ***	-0.067
Extractives	L					-0.703 *
content (EXTR)	F					0.167
(age 13)	I					-0.069

Discussion

Variation in and heritability of wood properties and growth

Two of the four chemical wood traits examined had significant variation at either the locality or family within locality level, indicating there is genetic variation

within *E. globulus*. Useful heritability estimates were also obtained for several traits despite their relatively large standard errors due to the small sample size. Both Klason lignin and acid-soluble lignin contents showed significant variation among localities, which suggested improvement could be made through locality selection. Surprisingly no locality differences were found for total lignin content, although this is consistent with the study of Miranda and Pereira (2002) using five trees of *E. globulus* from each of four provenances. Acid-soluble lignin content showed significant variation for families and the highest estimated narrow-sense (0.51) and family means (0.64) heritabilities. The only published narrow-sense heritability estimate for lignin traits in *E. globulus* is for total lignin content which was estimated to be very low at 0.09 ± 0.21 (Cotterill and Brolin, 1997). The moderate narrow-sense heritability estimate for total lignin content from the current study (0.29), together with a high family means heritability (0.50), suggest that lignin may be under stronger genetic control than previously thought. Supporting this is an estimate for the clonal heritability of lignin content in *E. globulus* of 0.83 from Gominho et al. (1997).

Extractives content was found to have a moderate narrow-sense heritability (0.35 ± 0.23), but no statistically significant differences were found between or within localities. Significant provenance effects for extractives content have been found previously in *E. globulus* (Washusen et al., 2001; Miranda and Pereira, 2002), suggesting provenance selection could be used to improve this trait. The lack of variation in the current study may be due to different provenances being used, or may be attributed to possible site by genotype interactions affecting this trait. Kube (2005) found strong genotype by site interactions for extractives among 434 *E. nitens* trees from 40 families grown over three sites, with heritability estimates found to vary between sites from low to very high, suggesting that the factors causing extractives production in some genotypes are very site specific. Miranda and Pereira (2002) found no site effects for extractives while the current study found no replicate effects in *E. globulus*.

The heritability of and the variation in basic density and diameter for *E. globulus* has been examined extensively (Lopez et al., 2002) and so will not be discussed in detail here. Basic density (age 13) did not have significant variation at the family or

locality level, although the trends in locality means (King Island low and the Strzelecki localities high) were consistent with previous studies that have reported significant differences (Dutkowski and Potts, 1999; Muneri and Raymond, 2000; Apiolaza et al., in press). This suggested that the small sample size and the use of only the outer part of the core reduced the power of the current study and therefore significance would generally be underestimated.

All of the trees used in this study had been cored previously which meant tissue had potential exposure to infection by wood decaying organisms. The West Ridgley site is also a wet site which has been found to be a factor leading to an increase in the incidence of decay (Mohammed et al., 2000). Localities differed significantly in the extent of decay with South King Island found to be particularly susceptible. This was the first evidence of genetic variation for decay resistance in *E. globulus*. The two main races of *E. globulus* that have been used for plantation growth in Australia, Strzelecki and King Island (Potts et al., 1999), were placed at either end of the range in decay as they have been previously for basic density (Dutkowski and Potts, 1999). The fast growing but low density King Island trees were originally grown for pulp production, however, Strzelecki and Western Otways became preferred (Potts et al., 1999). It appears that high basic density trees now selected in the breeding program may be more resistant to decay. Although the narrow-sense heritability for decay was estimated here as zero, a high family means heritability was obtained (0.50 ± 0.14). Narrow-sense heritability estimates in *E. nitens* have been found to vary between studies from 0.13 to 0.41 (White et al., 1999; Kube, 2005), and also between sites in a single study ranging from 0.04 to 0.63 (Kube, 2005). The successful exclusion of decay is likely the result of a number of traits including lignin and extractives contents, and therefore environmental and site influences are likely to be strong (Kube, 2005).

Correlations amongst wood properties

Phenotypic correlations indicate the presence of relationships between traits that may be due to a similar response to environmental conditions or to genetic associations. Genetic correlations are important for determining the potential for concurrent or independent selection of traits. Correlations between family means

(adjusted for locality differences) and between locality means, were used to give an indication of the genetic associations for this dataset. No study has yet identified the genetic correlations among the chemical wood traits (excluding pulp yield and cellulose content) and their correlated effects on the physical wood traits and growth in *E. globulus*.

Correlations amongst the chemical wood traits were often strong and as expected in terms of kraft pulping properties (Smook, 1992). A high pulp yield and cellulose content was associated with low extractives, Klason lignin and total lignin contents at both the phenotypic and genetic levels. This was consistent with the phenotypic correlations reported by Wallis et al. (1996) for 11 individuals of *E. globulus*. Miranda and Pereira (2001) examined 37 provenances of *E. globulus* and reported a similar correlation between pulp yield and extractives content at the provenance level, but not with total lignin content. No significant correlations were identified between acid-soluble lignin content and Klason lignin content, consistent with the findings of Miranda and Pereira (2001) who suggested differences in the lignin composition may be responsible. Lignin and extractives contents were generally positively correlated here and Ona et al. (1998) found similar relationships in a within-tree study of two *E. globulus* individuals. In *E. nitens* Kube and Raymond (2001) reported a very high negative genetic correlation between extractives and cellulose contents. These studies collectively suggest that selection for increased pulp yield or cellulose content are likely to result in a reduction in lignin and extractives contents, which are favourable responses for a pulpwood breeding objective.

The correlated effect of lignin on wood density is interesting as density is one of the main selection traits in the *E. globulus* breeding program. Basic density at ages 11 and 13 were significantly positively correlated at most levels ($r = 0.56$, $P < 0.01$ for families and $r = 0.65$, $P < 0.0001$ for individuals), and both were negatively correlated with lignin content at the family level. No other studies have looked at the relationship between lignin and basic density for larger sample sizes in *Eucalyptus*. However, a negative genetic correlation has also been found between density and lignin content in *Pinus pinaster* (Pot et al., 2002). It is therefore likely that favourable lignin profiles are being indirectly selected along with high basic

density. Similar to other studies in *E. globulus* (Ona et al., 1998; Miranda and Pereira, 2001) no apparent relationship was found between basic density and extractives content, although there are reports of positive associations in both *E. globulus* (Washusen et al., 2001) and *E. nitens* (Kube and Raymond, 2001).

Positive phenotypic and genetic correlations were found between microfibril angle and lignin content which is consistent with observations for coniferous wood (Saka, 1984). This relationship is thought to be due to the distribution of the microfibrils about their preferred orientation being large when the microfibril angle is large, therefore creating an imperfect alignment with more room for lignin deposition (Walker and Butterfield, 1995). These results suggest that a reduced microfibril angle (which gives the fibre a greater tensile strength and decreases its shrinkage [Bootle, 1983]), may be associated with improved lignin profiles for pulping.

Decay resistance is unlikely to become a major focus for selection in breeding programs for pulpwood, however, it is an important issue in the production of solid wood (Kube, 2005). Understanding the genetic relationships between decay and the chemical and physical wood traits, as well as growth, is therefore important. When examining relationships between decay incidence or extent with wood chemistry, it is important to distinguish between the chemistry found for normal healthy wood, and that found in diseased wood or in the reaction zone between healthy and diseased wood. It has been reported previously that the extractives and lignin contents are elevated in response to decay in eucalypts (Barry et al., 2001; Hawkins and Boudet, 2003), with the extractives content found to be six times greater in the reaction zone compared to healthy sapwood (Barry et al., 2000). Only negative locality level correlations were found in the current study between the extent of decay and both extractives and lignin contents. No correlations have been found between extent of decay and extractives content in *E. nitens* (Kube, 2005), however, a negative relationship has been found in *E. delegatensis* (Wong et al., 1983). Together these studies indicate that increases in extractives and lignin contents may only occur for diseased wood or in the reaction zone (both of which were removed in the current study), and the surrounding, healthy wood has a normal extractives level.

A negative correlation between the extent of decay and basic density (age 11) was observed at the locality level, which seemed to be the result of one locality (South King Island) that appeared to be particularly susceptible to decay and is known for its low basic density (Dutkowski and Potts, 1999). However, the extent of decay showed significant phenotypic correlations with basic density (age 11, negative), even with the South King Island locality removed from the analysis. Similar negative correlations have also been found in *E. delegatensis* and *E. grandis* (Nelson and Heather, 1972; Wong et al., 1983). It has been proposed that lower density wood has wider cell lumina, and therefore a larger surface area is exposed to the enzymes of decay micro-organisms, and also the water and air content in the wood may be at a level that promotes fungal growth (Southam and Ehrlich, 1943). A positive genetic correlation between the extent of decay and mean fibre length was also found, and may support this idea. The lack of a significant correlation between the extent of decay and basic density at age 13, may be because the decayed area of the core was removed prior to basic density measurements and only partial cores were used. The age 11 measures of wood density may have been taken before the formation of the decay and may be more indicative of wood susceptibility to decay.

The combination of the chemical wood properties with the physical wood properties of Apiolaza et al. (in press), allows a primary analysis of the genetic variation of the most important traits associated with pulp production, and how they are correlated with one another. This is the first study incorporating such a large number of traits for *E. globulus*, although the results must be treated with some caution due to the small sample size. The results indicate that when selecting for the current breeding objective traits of high basic density and pulp yield (Greaves et al., 1997), other traits beneficial to the pulping process may concurrently be selected, including low lignin and extractives contents, and a high cellulose content, as well as improved fibre properties. Selection for high basic density may also result in increased resistance to decay. Growth may be selected for independently of most of the chemical wood properties and decay resistance.

CHAPTER 6

The effect of a single amino acid substitution in a lignin biosynthesis enzyme on wood properties in *Eucalyptus globulus*

This chapter was published in part as: Poke F.S., Raymond C.A., Reid J.B., Vaillancourt R.E. (2004) The effect of a single amino acid substitution in a lignin biosynthesis enzyme on wood properties in *Eucalyptus globulus*. *In* Proceedings of IUFRO conference "*Eucalyptus* in a changing world", edited by N.M.G. Borralho, J.S. Pereira, C. Marques, J. Coutinho, M. Madeira, M. Tome, pp 388-394. Aveiro, Portugal, Oct 11-15. Note: data and figures in this reference have been modified to take into account more recent genotyping and further chemical analyses.

CHAPTER 7

The impact of intragenic recombination on phylogenetic reconstruction at the sectional level in *Eucalyptus* when using a single copy nuclear gene (cinnamoyl CoA reductase)

This chapter has been published as: Poke F.S., Martin D.P., Steane D.A., Vaillancourt R.E., Reid J.B. (2006) The impact of intragenic recombination on phylogenetic reconstruction at the sectional level in *Eucalyptus* when using a single copy nuclear gene (cinnamoyl CoA reductase). Molecular Phylogenetics and Evolution 39(1): 160-170.

Introduction

Eucalyptus is a genus of vital importance to the Australian environment and to the forestry industry world-wide, being grown for pulpwood and solid timber products throughout Australia, South America, Asia, Africa and Southern Europe. Understanding evolutionary relationships among eucalypts is difficult because of the diversity of the genus (includes approximately 700 species) and the propensity for interspecific hybridisation (Griffin et al., 1988). As a result, there has been much debate about the taxonomy and phylogeny of the genus. The most recent taxonomy of the eucalypts was done by Brooker (2000), and was based solely on morphological data. DNA studies also have been used extensively to elucidate the relationships among the eucalypts (Udovicic et al., 1995; Chappill and Ladiges, 1996; Sale et al., 1996; Jackson et al., 1999; McKinnon et al., 1999; Steane et al., 1999; Udovicic and Ladiges, 2000; Steane et al., 2002; Whittock et al., 2003), frequently producing phylogenies that are at odds with morphology-based classifications (Brooker, 2000).

Systematic studies based on DNA have been conducted using chloroplast DNA (cpDNA) or nuclear encoded ribosomal DNA (nrDNA). cpDNA is easy to work with because it is haploid, it has a high copy number and it is unaffected by recombination (Small et al., 2004). However, there are limits to what can be achieved with cpDNA analyses. cpDNA is maternally inherited in eucalypts (Byrne et al., 1993; McKinnon et al., 2001b) and therefore represents only half of the parentage (Sang, 2002), it has a low interspecific sequence diversity and is therefore most useful at higher taxonomic levels such as intergenus and interfamily levels (Sang, 2002; Small et al., 2004), and identical haplotypes are found in different species, attributable partly to interspecific hybridisation (McKinnon et al., 1999; McKinnon et al., 2001a; McKinnon et al., 2004). These limitations considerably impede cpDNA-based phylogenetic reconstructions of species relationships. nrDNA has subsequently been explored to overcome some of these limitations and to provide an independent assessment of phylogenies. Like cpDNA, nrDNA has a high copy number, but evolves at a faster rate than cpDNA and therefore has more sequence diversity (Small et al., 2004). nrDNA comprises genes that are useful for higher-level taxonomic relationships, and intergenic spacers that are useful for lower taxonomic levels (Small et al., 2004). Most commonly the 5S nrDNA spacer and the internal transcribed spacer (ITS) have been used for phylogenetic studies of the eucalypts (Udovicic et al., 1995; Steane et al., 1999, 2002), and in general they have been quite successful in elucidating evolutionary relationships. Steane et al. (2002), in the most comprehensive phylogenetic analysis of *Eucalyptus* to date, found that ITS sequence data gave good resolution at higher taxonomic levels (e.g. between subgenera), but lower level relationships (e.g. at the sectional level) were more difficult to resolve.

More recently, the utility of low copy number nuclear genes (LCNG) in phylogenetic analyses has been explored (see review by Sang, 2002; Small et al., 2004). This approach has both distinct advantages and limitations. LCNG generally evolve faster than either cpDNA or nrDNA. The rate of sequence evolution differs between genes and regions of genes (exons, introns and untranslated regions), and this may assist taxonomic resolution at various levels (Sang, 2002; Small et al., 2004). LCNG are biparentally inherited and so, unlike cpDNA, provide information from both the maternal and the paternal lineages. Although nrDNA is biparentally

inherited, the isolation of individual parental sequences is difficult due to the high gene copy numbers involved (Small et al., 2004). The availability of multiple unlinked loci in the nuclear genome may provide a useful resource for independent phylogenetic analysis when incongruence is found (Sang, 2002). LCNG may, however, have more complex genetic architectures and evolutionary dynamics than cpDNA and nrDNA, which may limit their use (Small et al., 2004). Some of these complications include recombination, concerted evolution, selection and heterozygosity (see review by Sang, 2002; Zhang and Hewitt, 2003; Small et al., 2004).

One of the major limitations of using nuclear genes for phylogenetic reconstruction is recombination (reviewed in Posada et al., 2002). Recombination is the reciprocal exchange of genetic material between two homologous chromosomes during meiosis (Griffiths et al., 2005). Its frequency varies between loci, is influenced by chromosomal location (regions near centromeres and telomeres show little recombination) and sequence structure, and has been found to occur within single genes (Smouse, 2000; Zhang and Hewitt, 2003). Intragenic recombination generates alleles that are chimeric between parental alleles (Small et al., 2004) and therefore when it has occurred, the evolutionary history of a set of sequences forms a group of contradictory phylogenetic trees rather than a single tree (Zhang and Hewitt, 2003). When using nuclear genes for phylogeny reconstruction, avoiding regions that have been influenced by recombination may be difficult, because a positive correlation exists between recombination rate and the level of sequence polymorphism (Zhang and Hewitt, 2003). Therefore genomic regions with low recombination rates may not have enough sequence variation for phylogenetic analysis. The likely strategy for overcoming this problem is to analyse recombination events in the dataset and incorporate them into the models of evolution (Zhang and Hewitt, 2003).

Several methods have been developed for detecting the presence of recombination, identifying the parental and recombinant individuals, and approximating the positions of recombination break-points (reviewed in Posada and Crandall, 2001). These generally involve four different detection strategies: distance methods that look for inversions of distance patterns among the sequences; phylogenetic methods

that compare the branching patterns of adjacent sequences; compatibility methods that partition phylogenetic incongruence site by site; and substitution distribution methods that look for a significant clustering of nucleotide substitutions or an expected statistical distribution (Posada and Crandall, 2001). Detection of recombination by more than one method should always be attempted before conclusions are drawn about the presence of recombination. The performance of these methods for detecting recombination has been evaluated using both computer simulations and empirical data, and has been found to depend on the amount of recombination present, genetic diversity and evolutionary rate variation among sites (Maynard Smith, 1999; Posada and Crandall, 2001; Wiuf et al., 2001; Posada, 2002). In general, substitution methods have been found to be more powerful than phylogenetic methods (Posada and Crandall, 2001; Posada, 2002).

This study explored the utility of a single copy nuclear gene, cinnamoyl CoA reductase (*CCR*) (Lacombe et al., 1997; Poke et al., 2003), for resolving interspecific phylogenetic relationships within *Eucalyptus*. *CCR* is a lignin biosynthesis gene that displays a high level of sequence diversity within *Eucalyptus globulus*, particularly within introns (Poke et al., 2003; Appendix 1). Previous phylogenetic analysis of *Eucalyptus* using the ITS region of the nrDNA found that sections *Exsertaria* and *Latoangulatae* formed a monophyletic group, but that differences between the two sections could not be resolved (Steane et al., 2002). This study aimed to use a single copy nuclear gene to determine whether or not these sections could be resolved as monophyletic groups. The occurrence of recombination within this dataset and its effect on the phylogenetic outputs were critical elements of this analysis.

Materials and Methods

Genetic material

Leaf material from 23 *Eucalyptus* species (a total of 29 samples; Table 1) in sections *Latoangulatae* (SL), *Exsertaria* (SE) and *Maidenaria* (SM) of subgenus *Symphyomyrtus* was collected from the Currency Creek Arboretum, South Australia (http://www.dn.com.au), and from existing frozen stocks used in previous studies

(Sale et al., 1993; Bundock et al., 2000; Steane et al., 2002; Poke et al., 2003; Rathbone, 2003). *Symphyomyrtus* is one of 13 subgenera of *Eucalyptus* and is composed of 15 sections (Brooker, 2000), of which three are represented here. SM has previously been identified as a monophyletic group closely related to SE and SL (Steane et al., 2002), and so was included here as an outgroup. DNA was extracted from 50 mg of leaf material using a modified protocol of Doyle and Doyle (1990). A fragment of the *CCR* gene was PCR amplified from each sample in a Corbett Research Palm Cycler (NSW, Australia) using a 20 μL reaction volume comprising 2 μL BD Biosciences 10X Advantage 2 PCR buffer (BD Biosciences CA, USA), 0.5 μL dNTPs (10 mM stock solution), 0.5 μL BD Biosciences Advantage 2 polymerase mix, 1 μL each of primer 420 (5' TTA TGT GCG TGT AGA CGA CCC GAA GAA 3') and 3198 (Poke et al., 2003), 1 μL DNA and 14 μL dH_20. A two-step PCR cycle was used involving an initial denaturation of 4 min at 94°C, 35 cycles of 15 sec at 94°C and 3 min at 70°C, followed by a final extension of 12 min at 72°C.

PCR products were cleaned using a QIAGEN QIAquick PCR purification kit (QIAGEN Pty Ltd Vic, Australia). A portion of the *CCR* gene (approximately 44% of the gene) including the complete intron 4 and parts of exons 4 and 5, was sequenced. This region was chosen because it incorporated the longest intron within the gene, and the exons contained no known important functional motifs such as the catalytic site or the cofactor binding site, and consequently may be under less selective pressure (Lacombe et al., 1997; Pichon et al., 1998). Cycle sequencing reactions were performed on a Corbett Research Palm Cycler according to the Beckman Coulter protocol, modified for a 10 μL reaction volume and using a final concentration of 1 μM sequencing primer and 50-100 fmol of PCR product (Poke et al., 2003). PCR products were sequenced in the forward and reverse directions using a set of internal sequencing primers 1588 (Poke et al., 2003), 2043 (5' AAA ACT TAG ATA GAT AG 3'; G.E. McKinnon, personal communication), 2090, 2531 (Poke et al., 2003), 2639 (5' GGG GAG GGT AGG TGA GGA TA 3') and 3085 (5' TGC ACC GTG ATG GAT CTA AG 3'). Sequencing was done on a Beckman Coulter CEQ 2000 automated sequencer (Beckman Coulter Inc CA, USA). Sequences were analysed and aligned using Sequencher software (Gene

Codes Corporation MI, USA) and manually checked for incorrect base calls, heterozygosity, polymorphisms and insertions/deletions (indels).

Table 1. Specimen details for the *Eucalyptus* samples from sections *Latoangulatae*, *Exsertaria* and *Maidenaria* used in the phylogenetic and recombination analyses of the *CCR* gene. Samples with two GenBank numbers have separate 5' and 3' sequence fragments, due to the presence of a long microsatellite in the middle of those two regions, which was difficult to sequence. CCA refers to the Currency Creek Arboretum, South Australia.

Section	Species	Location	Source	GenBank Accession Number(s)
Exsertaria	*alba*	Waite Arboretum, South Australia	Sale et al., 1993	DQ084782
Exsertaria	*blakelyi*	Bingara, New South Wales	CCA	DQ084783
Exsertaria	*brassiana*	Kypiano, Papua New Guinea	Steane et al., 2002	DQ084786
Exsertaria	*chloroclada*	Warwick, Queensland	CCA	DQ084787
Exsertaria	*cupularis*	Halls Creek, Western Australia	CCA	DQ084788, DQ084789
Exsertaria	*flindersii*	Flinders Ranges, South Australia	CCA	DQ084791, DQ084792
Exsertaria	*glaucina*	Casino, New South Wales	Steane et al , 2002	DQ084793, DQ084794
Exsertaria	*hallii*	Goodwood, Queensland	Steane et al., 2002	DQ084801, DQ084802
Exsertaria	*infera*	Warwick, Queensland	CCA	DQ084803, DQ084804
Exsertaria	*nudicaulis*	Mt Isa, Queensland	CCA	DQ084813
Exsertaria	*rudis* ssp. *rudis*	Pallinup River, Western Australia	CCA	DQ084816, DQ084817
Exsertaria	*vicina*	Cobar, New South Wales	CCA	DQ084822
Exsertaria	aff *vicina*	Byngnano Range, New South Wales	CCA	DQ084807, DQ084808
Latoangulatae	*botryoides* (1)	Australian National Botanic Gardens, Australian Capital Territory	Sale et al , 1993	DQ084784
Latoangulatae	*botryoides* (2)	Western Australia	Sale et al., 1993	DQ084785
Latoangulatae	*deanei*	Wisemans Ferry, New South Wales	CCA	DQ084790
Latoangulatae	*grandis* (1)	Tinaroo, Queensland	Wild collection	DQ084798
Latoangulatae	*grandis* (2)	Black Hill Flora Centre, Athlestone, South Australia	Sale et al., 1993	DQ084800
Latoangulatae	*grandis* (3)	Paluma, Queensland	Steane et al , 2002	DQ084799
Latoangulatae	*major*	Helidon, Queensland	CCA	DQ084805, DQ084806
Latoangulatae	*notabilis*	Blue Mountains, New South Wales	CCA	DQ084811, DQ084812
Latoangulatae	*punctata*	Nowra, New South Wales	CCA	DQ084815
Latoangulatae	*saligna* (1)	Grafton, New South Wales	CCA	DQ084818
Latoangulatae	*saligna* (2)	Blackdown Tableland, Queensland	CCA	DQ084819
Latoangulatae	*scias* ssp. *apoda*	Tenterfield, New South Wales	CCA	DQ084820, DQ084821
Maidenaria	*globulus* (1)	King Island, Tasmania	Bundock et al., 2000	DQ084797
Maidenaria	*globulus* (2)	King Island, Tasmania	Poke et al., 2003	DQ084795, DQ084796
Maidenaria	*nitens*	Rubicon, Victoria	Steane et al., 2002	DQ084809, DQ084810
Maidenaria	*perriniana*	Howqua Gap, Victoria	Rathbone, 2003	DQ084814

Phylogenetic analysis

Phylogenetic analysis of the 29 *Eucalyptus CCR* sequences was done using PAUP 4.0b10 (Swofford, 1999). Maximum parsimony analysis was performed using the branch and bound algorithm with all uninformative characters excluded. Analyses were conducted either with indel sequence included or with indel sequence excluded and recoded as described in Table 2. Strict consensus trees were constructed with *E. nitens* (SM) used as the outgroup, based on previous

phylogenetic analysis of *Eucalyptus* (Steane et al., 2002). Bootstrap support was estimated using the branch and bound algorithm with 10,000 bootstrap replicates.

Recombination analysis

Detection of potential recombinant sequences, identification of likely parental sequences, and localization of possible recombination breakpoints were carried out using the RDP (Martin and Rybicki, 2000), GENECONV (Padidam et al., 1999), BOOTSCAN (Martin et al., 2005a), and MAXIMUM CHI SQUARE (Maynard Smith, 1992) methods as implemented in RDP2 (Martin et al., 2005b). An alignment of the 26 unique *CCR* sequences was scanned using these methods with the following settings: RDP – six nucleotide windows with no reference sequences used; GENECONV – sequence triplets scanned using a g-scale of 1; BOOTSCAN – 100 nucleotide windows with a 20 nucleotide step distance scanned, using a 70% bootstrap cutoff and 100 bootstrap replicates with Jukes Cantor distances and calculation of binomial P-values; and MAXIMUM CHI SQUARE – sequence triplets scanned using a fixed window size of 16 nucleotides with gaps stripped. Potential recombination signals were identified as those detectable by two or more methods using an uncorrected P-value cutoff of 0.01.

The P-value cutoff of 0.01 was quite lenient and false positive signals were expected. To determine the expected number of false positive signals a permutation test was used. The positions of columns in the *CCR* alignment were randomly shuffled 100 times and the shuffled alignments were analysed in the same way as above. The false positive rate was approximated as the average number of unique recombination-like signals detected in the shuffled alignments. The global probability of recombination using the combined RDP, GENECONV, BOOTSCAN, and MAXIMUM CHI SQUARE analyses (i.e. the probability that at least one real recombination event was detectable within the analysed alignment) was the proportion of shuffled alignments in which the number of unique detectable recombination events equalled or was greater than the number detected in the unshuffled alignment.

Results

Sequence analysis

Complete sequence was obtained for the 29 *Eucalyptus* samples from sections *Latoangulatae*, *Exsertaria* and *Maidenaria* for the *CCR* fragment (GenBank accession nos. DQ084782–DQ084822; Table 1). The sequences formed an alignment of 1418 base pairs in length, of which base pairs 1-337 corresponded to the 3' end of exon 4, 338-1299 to intron 4 and 1300-1418 to the 5' end of exon 5. These sequence positions do not take into consideration a region of approximately 47 base pairs, which was removed between positions 698 and 699 due to the presence of a variable length CT repeat microsatellite. This microsatellite region was excluded from further phylogenetic and recombination analyses because it made alignment of the sequences difficult and may have contained sequencing errors due to the difficulty associated with sequencing highly repetitive DNA. No polymorphisms were detected in the chromatograms which suggested that not more than one copy of the *CCR* gene was being sequenced in each sample. Eight indels ranging from 2 to 35 base pairs in length were found and are described in Table 2. A total of 184 polymorphic sites (including substitutions and one base pair indels, and excluding large indels) were found in the *CCR* fragment which equated to one polymorphic site every seven base pairs of the alignment (Table 3). The intron region (one polymorphic site every six base pairs) was nearly two times as polymorphic as the total exon region (one polymorphic site every eleven base pairs) (Table 3). The exon regions showed varying rates of sequence diversity with exon 4 (one polymorphic site every ten base pairs) twice as polymorphic as exon 5 (one polymorphic site every twenty base pairs) (Table 3). Of the 184 polymorphisms, 61 were potentially informative for phylogenetic analysis.

Table 2. Insertions/deletions (indels) found in intron 4 of the *CCR* gene fragment from 29 *Eucalyptus* samples of sections *Latoangulatae*, *Exsertaria* and *Maidenaria*. Deletion of sequence in each position was coded as zero; each insertion was coded as shown.

Character	Sequence position	Character states and indel sequence
1	440-443	1 ATAG
2	543-545	1 TGA
		2 TGR
		3 CGA
3	561-595	1 AATAGAATTACTCAAGCTGATGTGGTCCAAGCGTC
		2 AATAGAATTACTGAAGCTGATGTGGTCCAAGCGTC
4	855-857	1 AGG
5	1016-1017	1 TT
6	1029-1031	1 GAG
7	1154-1155	1 AT
8	1231-1246	1 CATGTGATGTGTCGAG

Table 3. Sequence polymorphisms (including substitutions and one base pair insertion/deletions) found in the *CCR* gene fragment for 29 *Eucalyptus* samples from sections *Latoangulatae*, *Exsertaria* and *Maidenaria*. The size of a region was calculated from the aligned sequences and excluded the insertion/deletion sequences described in Table 2.

Sequence	Size in base pairs	Number of polymorphic sites	Polymorphic sites per base pairs
3' exon 4	337	35	1/10
5' exon 5	119	6	1/20
Total exon sequence	456	41	1/11
Total intron sequence	894	143	1/6
Total sequence	1350	184	1/7

Phylogenetic analysis

Phylogenetic tree construction by maximum parsimony found 12 equally most parsimonious trees of length 118 (CI = 0.542, RI = 0.831). The strict consensus of these trees is shown in Figure 1. Analysis with the indels scored as binary characters produced an identical topology to that obtained when indel sequences were retained (12 trees of length 127, CI = 0.551, RI = 0.829). None of the three sections of *Eucalyptus* was monophyletic. While all SE species formed a single clade, there was low bootstrap support for this clade (53%), which also included a single SL species (*E. major*). SL and SM were polyphyletic, with *E. globulus* 1

(SM) grouping with the SL species, and *E. deanei* (SL) grouping with the SM species (Figure 1). Furthermore, even if we ignored the anomalous positions of these two species, SL would be paraphyletic. When two regions of the sequence (i.e. downstream [1-698] and upstream [699-1418] of the excluded microsatellite region) were analysed independently, the strict consensus trees showed little resolution (results not shown). The two regions of the sequence yielded different topologies among SL species. This suggested that different regions of the gene had different evolutionary histories signifying that recombination may have occurred.

Recombination analysis

Simulation studies have demonstrated some of the difficulties inherent in detecting recombination amongst sequences with low diversity (Posada and Crandall, 2001). These studies have indicated that methods employing "global" permutation tests for the presence of recombination generally outperform those using "local" Bonferroni corrected statistical tests. Unfortunately, whereas local tests give information on individual statistically probable events such as breakpoint positions and which sequences are recombinant, global tests generally only indicate whether or not there is significant evidence of recombination in an alignment. The relatively low degree of sequence diversity in the *CCR* genes that were analysed meant that, even if evidence of recombination was present, it would be very difficult to detect and characterize individual recombination events using even the most powerful local tests designed for this purpose. A global modification of four of the most powerful local recombination tests available (Posada and Crandall, 2001; Posada, 2002; Martin et al., 2005a,b) was devised that involved counting and comparing the number of unique recombination events detectable in unshuffled and shuffled alignments.

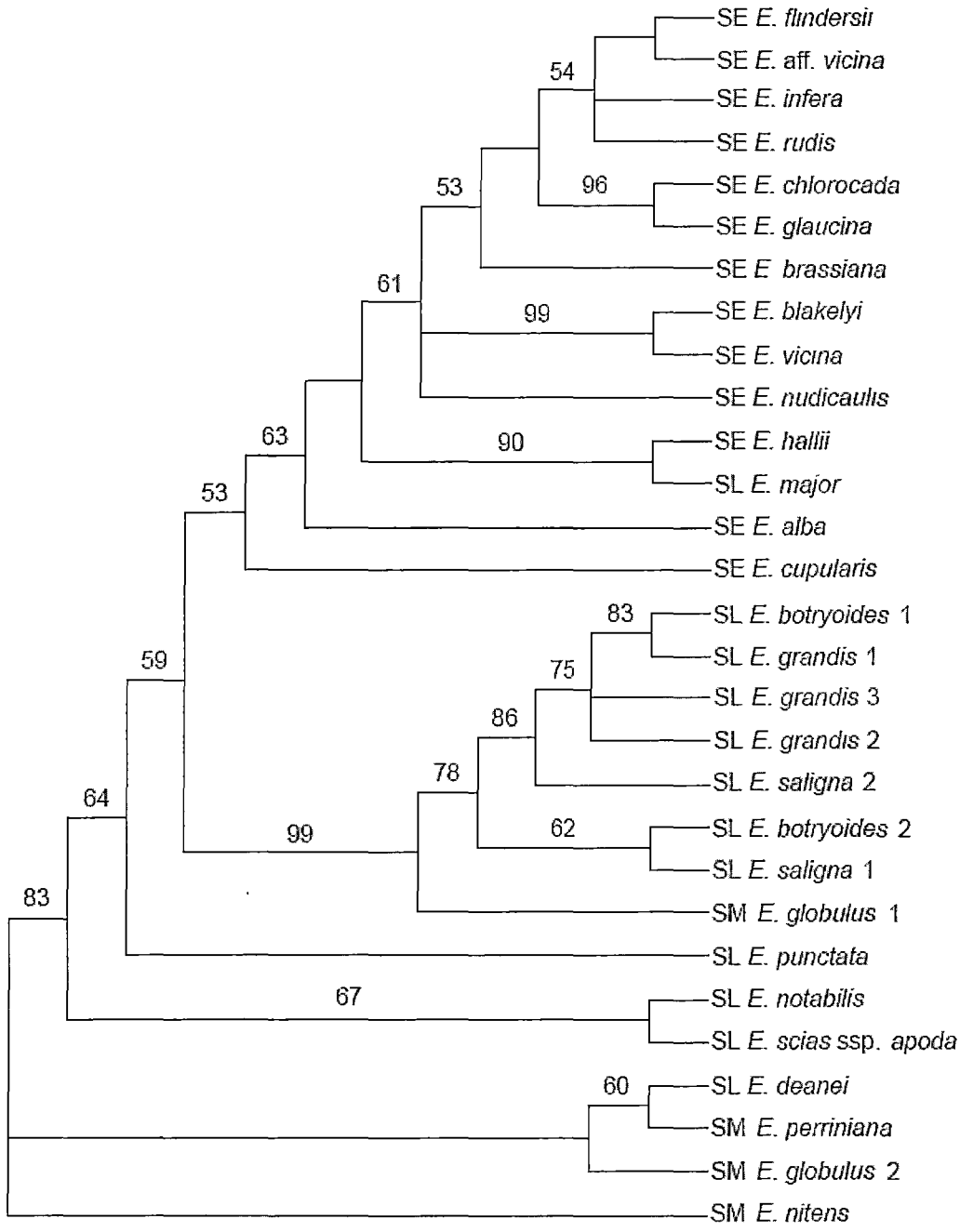

Figure 1. Strict consensus of 12 trees from the phylogenetic analysis of species in sections *Exsertaria* (SE), *Latoangulatae* (SL) and *Maidenaria* (SM) using the *CCR* gene fragment. Phylogenetic trees were constructed using the branch and bound algorithm in PAUP. Numbers above branches represent bootstrap percentages after 10,000 bootstrap replicates.

The permutation test using the RDP, GENECONV, BOOTSCAN, and MAXIMUM CHI SQUARE recombination detection methods indicated probabilities of 0.14, 0.05, 0.02 and < 0.01 respectively that the *CCR* alignment contained more evidence of individual recombination events than did the shuffled alignments. When only potential events detectable by two or more of the four methods were considered, the probability that the unshuffled alignment contained more evidence of individual recombination events than the shuffled alignments was < 0.01 (i.e. no shuffled alignment had more detectable events than did the unshuffled alignment). Of the 21 unique potential events detected in the unshuffled alignment, only eleven were detectable with two or more methods. Whereas the average number of unique events detectable by two or more methods in the 100 shuffled alignments was 4.5, the greatest number was nine. This indicated that approximately five (41%) of the eleven potential events detected in the unshuffled alignment were likely to be false positives. With only two exceptions, determining which of the eleven potential events were real was not possible.

The two exceptional events were detected by the MAXIMUM CHI SQUARE method with an associated Chi square P-value that was unmatched by any false positive event detected in any of the shuffled alignments. In the first event, SL species *E. botryoides* 1, *E. grandis* 1, *E. grandis* 2, *E. saligna* 1, *E. saligna* 2 were likely to be descendents of the recombinant progeny of species with sequences similar to *E. nudicaulis* (SE) and *E. scias* ssp. *apoda* (SL) (Figure 2). The recombinant region occurred between nucleotide positions 969 and 1211 (Figure 2). The second event identified the *E. globulus* 1 (SM) sequence to be a recombinant progeny of parental sequences similar to *E. grandis* 1 (SL) with the second parent unknown (outside the dataset), but was most similar to *E. scias* ssp. *apoda* (SL) (Figure 3). The recombinant region occurred between nucleotide positions 774 and 1211 (Figure 3). Phylogenetic evidence was obtained for the first event following construction of neighbour joining trees using PHYLIP (Felsenstein, 1989), with Kimura-2 parameter distances and 1000 bootstrap replicates, with different portions of *CCR* alignment (Figure 4).

Examining the breakpoint positions of the eleven potential events detected by two or more recombination detection methods in the unshuffled alignment indicated the

positions of two potential recombination hotspots in the *CCR* gene fragment (Figure 5). These occurred between alignment positions 230-451 and 890-1125. The former of these regions corresponded to the exon 4/intron 4 boundary. The latter region occurred in the last half of intron 4 which was associated with the presence of several indels (Table 2).

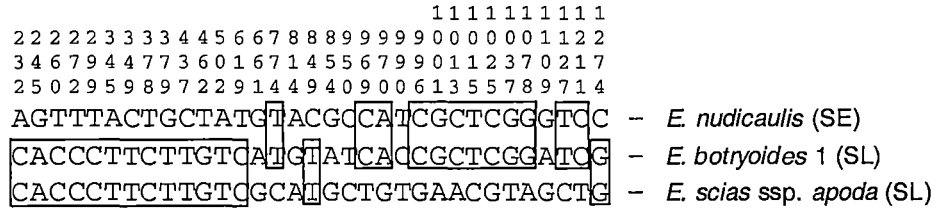

Figure 2. Representation of the recombination event in *CCR* depicting *E. botryoides* 1 (*Latoangulatae*, SL) as the recombinant progeny of the parental sequences *E. nudicaulis* (*Exsertaria*, SE) and *E. scias* ssp. *apoda* (SL). Only the variable sites between the three sequences are shown for the recombinant region of the *CCR* gene fragment. Numbers above the nucleotide alignments refer to base pair positions (read vertically). The recombinant *E. botryoides* 1 is very similar to *E. scias* ssp. *apoda* from nucleotide positions 1-968 and 1212-1418. However, it is more similar to *E. nudicaulis* at positions 969-1211. Sequences closely related to *E. botryoides* 1 (*E. grandis* 1, *E. grandis* 2, *E. saligna* 1 and *E. saligna* 2) were also identified as recombinants.

Figure 3. Representation of the recombination event in *CCR* where *E. globulus* 1 (*Maidenaria*, SM) was identified as the recombinant progeny of parental sequences similar to *E. grandis* 1 (*Latoangulatae*, SL) and an unknown parent (outside the dataset) that was most similar to *E. scias* ssp. *apoda* (SL). Only the variable sites between the three sequences are shown for the region of *CCR* at which recombination is thought to have occurred. Numbers above the nucleotides refer to base pair positions (read vertically). The sequence of *E. globulus 1* is most like *E. grandis* 1 in the recombinant region between positions 774-1211, however, in the rest of the sequence (1-773 and 1212-1418) it is more similar to *E. scias* ssp. *apoda.*

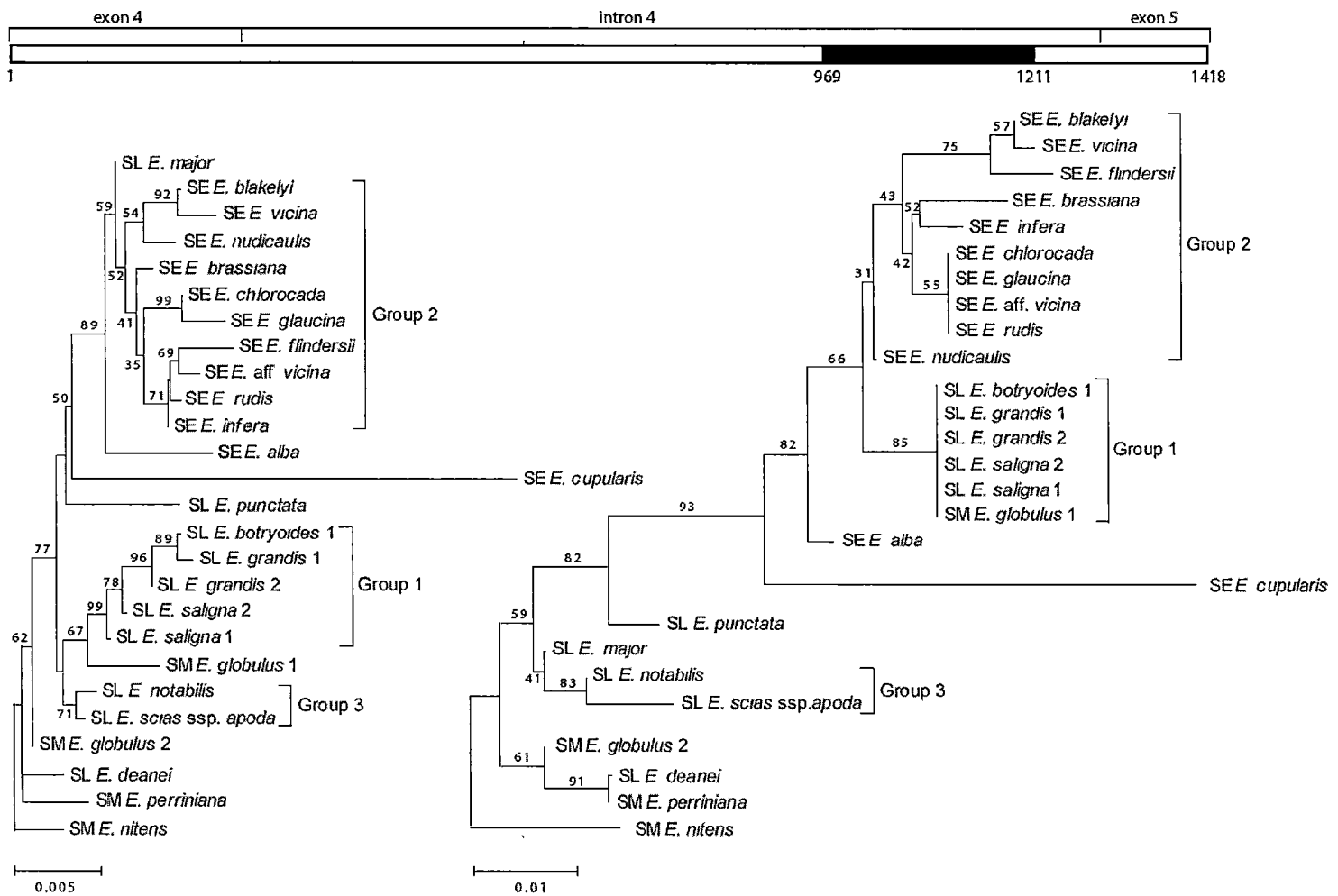
exon 4
intron 4
exon 5
1
969
1211
1418
SL E. major
SE E. blakelyi
SE E. vicina
SE E. nudicaulis
SE E. brassiana
SE E. chlorocada
SE E. glaucina
SE E. flindersii
SE E. aff vicina
SE E. rudis
SE E. infera
SE E. alba
Group 2
SE E. cupularis
SL E. punctata
SL E. botryoides 1
SL E. grandis 1
SL E. grandis 2
SL E. saligna 2
SL E. saligna 1
SM E. globulus 1
Group 1
SL E. notabilis
SL E. scias ssp. apoda
Group 3
SM E. globulus 2
SL E. deanei
SM E. perriniana
SM E. nitens
0.005
0.01

Figure 4 (previous page). Phylogenetic evidence for a potential recombination event in *CCR* of *Eucalyptus* species from sections *Maidenaria* (SM), *Latoangulatae* (SL) and *Exsertaria* (SE). Neighbour joining trees were constructed using PHYLIP with Kimura-2 parameter distances and 1000 bootstrap replicates. The tree on the left used nucleotide sequences depicted in white in the schematic representation of the alignment above the trees (positions 1–968 concatenated with positions 1212–1418). The tree on the right used nucleotide sequences depicted in black in the schematic representation of the alignment (positions 969–1211). Numbers above the branches represent bootstrap percentages. The common ancestor of Group 1 (SL *E. botryoides* 1, SL *E. grandis* 1, SL *E. grandis* 2, SL *E. saligna* 1 and SL *E. saligna* 2) was a recombinant sequence, and sequences related to the parental sequences involved in the recombination event are indicated by Group 2 (SE *E. nudicaulis*) and Group 3 (SL *E. scias* ssp. *apoda*). A second recombination event occurred where SM *E. globulus* 1 was identified as a recombinant sequence, for which the most likely parents were SL *E. grandis* 1 (Group 1) and a sequence similar to that of SL *E. scias* ssp. *apoda* (Group 3). This caused the placement of *E. globulus* 1 in Group 1 in the tree on the right.

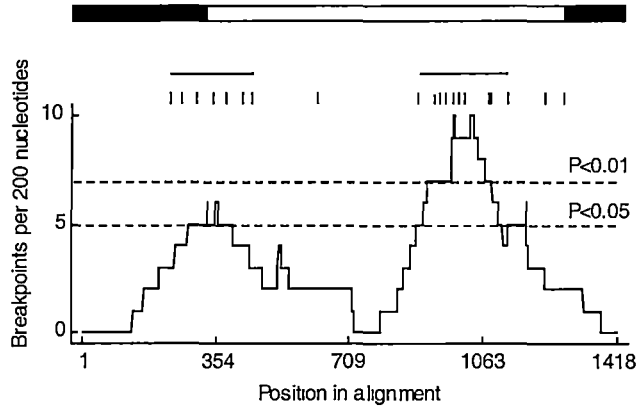

Figure 5. Recombination hotspots within the alignment of 26 *CCR* gene fragments for species from sections *Maidenaria*, *Latoangulatae* and *Exsertaria* of *Eucalyptus*. Detectable breakpoint positions (indicated by small vertical lines at the top of the graph) suggest the location of two potential recombination hotspots (indicated by horizontal lines above the graph). The bar at the top of the graph represents the *CCR* fragment: black represents exon regions and white depicts the intron. A 200 nucleotide window was moved along the alignment one nucleotide at a time and the number of breakpoints detected within the window region were counted and plotted. The upper and lower broken lines represent confidence levels; anything above those lines represents significant clustering based on a permutation test.

Discussion

The single copy status of *CCR* described previously for *Eucalyptus* (Lacombe et al., 1997; Poke et al., 2003: McKinnon et al., in press) was also indicated for the species examined here, with only one PCR amplification product observed and clearly only one gene sequence obtained per sample. Interspecific sequence variation in *CCR* was found to be high for *Eucalyptus*. The total gene fragment had on average one polymorphic site every seven base pairs (Table 3). Intron four was nearly twice as polymorphic as the exon region analysed, with one polymorphic site every six and eleven base pairs, respectively (Table 3). As might be expected, an evaluation of intraspecific sequence variation in *CCR* found much lower levels (Poke et al., 2003). Within *E. globulus*, one polymorphic site was found every 48 base pairs for the exons and every 33 base pairs for the introns (Poke et al., 2003), which is approximately one quarter and one sixth of the polymorphism described here, respectively. However, both studies identified similar ratios of sequence variation between the introns and the exons. The higher level of sequence variation in *CCR* observed between species (compared to within species), suggests that this gene may be useful for resolving interspecific phylogenetic relationships.

Phylogenetic analysis of the *Eucalyptus* species in sections *Latoangulatae*, *Exsertaria* and *Maidenaria* using the single copy gene, *CCR*, showed some consistency with taxonomic classification (Brooker, 2000). SE species formed a single clade, although this clade had low bootstrap support (53%) and included one SL species. SL and SM were polyphyletic. Three species (*E. major*, *E. deanei* and *E. globulus* 1) appeared in anomalous positions within the phylogeny: *E. major* (SL) occurred in the clade of SE species, an otherwise monophyletic group (90% bootstrap support for *E. major* and *E. halli* being sister taxa); *E. globulus* 1 (SM) formed a clade with species from SL (99% bootstrap support); and *E. deanei* formed a clade with SM taxa (60% bootstrap support that *E. deanei* and *E. perriniana* were sister taxa). Even with the exclusion of *E. major*, *E. deanei* and *E. globulus* 1, SL would be paraphyletic. While the phylogeny obtained using *CCR* did not necessarily correspond to the taxonomic classification of these species, high bootstrap supports indicate some well-supported relationships between species pairs and the clade comprising SL species and *E. globulus* 1.

The clades observed here contrasted with the phylogenetic reconstruction obtained by Steane et al. (2002) in an analysis of the ITS region of the nrDNA that could not separate *Latoangulatae* and *Exsertaria* species, and showed *Maidenaria* to be monophyletic. The current study found more resolution between sections *Latoangulatae* and *Exsertaria* using *CCR* than that found by Steane et al. (2002), although these sections still did not form monophyletic groups consistent with the morphology-based classification (Brooker, 2000). Paraphyletic groupings occur often in phylogenetic analyses (Brummitt, 2002; Mast et al., 2005; Pfeil and Crisp, 2005), particularly in the presence of reticulate evolution (Sosef, 1997) and it may be that *Latoangulatae* is paraphyletic. LCNG have been shown to be useful for reconstructing phylogenies in plants at both high and low taxonomic levels (Sang, 2002) including between genera (Mason-Gamer et al., 1998; Wang et al., 2000), species (Small et al., 1998; Emshwiller and Doyle, 1999; Lewis and Doyle, 2001; Tank and Sang, 2001) and populations (Olsen and Schaal, 1999; Olsen, 2002), often when cpDNA or nrDNA can not. However, our results suggest that phylogenetic reconstruction of the eucalypt sections based on *CCR* may be confounded by intragenic recombination.

Analysis of different parts of the *CCR* sequence showed little resolution of the phylogeny, with movement of the *Latoangulatae* groups and *E. major*, *E. deanei* and *E. globulus* 1 between different trees. The lack of resolution in these phylogenetic trees and the creation of conflicting trees from separate regions of the sequence, collectively suggested that intragenic recombination may be occurring (Posada and Crandall, 2001; Zhang and Hewitt, 2003). Recombination analysis confirmed the presence of recombination in *CCR*, although the low diversity of the sequence made it difficult to determine exactly how much recombination had occurred. Eleven possible recombination events were identified with five of these likely to be false positives due to the large number of tests performed. With only two exceptions it was not possible to infer which of the potential recombination events identified were likely to be real and which were not. However, phylogenetic evidence was found for the two most probable recombination events involving the three *Eucalyptus* sections. The first event indicated that SL species *E. botryoides*, *E. grandis* and *E. saligna* were descendents of the recombinant progeny of *E.*

nudicaulis -like (SE) and *E. scias* ssp. *apoda* -like (SL) species (Figure 2, Figure 4). This event would account for the movement of this group of SL species within the phylogenetic tree when different parts of the sequence were used for phylogenetic analysis (Figure 4). This recombination event is likely to represent an ancient event. Multiple species (Group 1) carry the recombinant allele, suggesting that recombination occurred prior to speciation in an ancestor of the group. The second event identified the sequence from *E. globulus* 1 (SM) as a recombinant, with one of the above recombinants, *E. grandis* 1, as a parent sequence, and an unknown sequence (outside the dataset) as the other parent (Figure 3). This second recombination event explained the inclusion of this SM species in a clade otherwise composed of SL species. The two *E. globulus* sequences in this dataset are quite different, with only one carrying the recombinant allele, suggesting that this recombination event may have occurred after speciation. However, the accumulation of mutations in the recombinant region of both *E. globulus* 1 and *E. grandis* 1 indicates that this event is unlikely to have occurred recently. This is supported by the findings of McKinnon et al. (in press) who identified two major *CCR* lineages in *E. globulus* that diverged shortly after the *Symphyomyrtus* sections diverged. Ancient recombination events in nuclear genes are likely to confound phylogenetic reconstruction of the eucalypts.

When using LCNG for phylogeny reconstruction, the analysis of recombination may be useful when incongruence is found, particularly when individual species occur in anomalous positions in the phylogenetic tree. If these samples are recombinants, alternative tree positions may be investigated based upon determination of the parental sequences, which may provide a better representation of the true evolutionary relationships. In this study, two *Latoangulatae* species (*E. deanei* and *E. major*) were included in clades composed of species from the other *Eucalyptus* sections, and these may be resolved in the future with elucidation of other recombination events. Although *E. major* was identified as a recombinant progeny of *E. vicina* (SE) and *E. punctata* (SL) by all four recombination detection methods (explaining the grouping of *E. major* with the SE species), it was not globally significant. No evidence was found for *E. deanei* being a recombinant using this part of the *CCR* sequence. However, this does not mean *E. deanei* is not a recombinant, because the only recombination events detectable were those that had

breakpoints within the analysed region of sequence. Simulation studies have shown that the phylogeny obtained using recombinant sequences can be very different from the true evolutionary history yielding those sequences (Posada and Crandall, 2002). When recombinant members of a species are found (such as *E. globulus* 1 in the current study), it may be beneficial to remove these from the analysis and use non-recombinant samples (*E. globulus* 2) to obtain a more accurate representation of phylogenetic relationships. This will be possible when recombination events have occurred after speciation. For ancient recombination events, where multiple species have the recombinant allele, this approach will be much more difficult, especially if taxonomic boundaries become blurred, as in this study. In this scenario it may be more useful to obtain an independent phylogeny using sequence data unaffected by recombination for comparison. Together, the results suggest that LCNG may be useful in resolving low level relationships amongst the eucalypts, but the gene, or region of the gene, must be chosen carefully, and recombination must be taken into consideration.

High frequencies of naturally occurring inter-sectional hybridisation have been reported for the sections *Latoangulatae*, *Exsertaria* and *Maidenaria*, particularly the former two (Griffin et al., 1988). Steane et al. (2002) suggested that the relationships found between the *Symphyomyrtus* sections using the ITS region correlated well with these reports, with the former two sections forming a clade and the latter a closely related monophyletic group. The occurrence of intragenic recombination within a nuclear gene involving three sections of eucalypts supports this idea. The incidence of hybridisation between these sections would facilitate the production of chimeric alleles between parents of two different sections through recombination. Therefore, phylogenetic reconstruction would be difficult for these sections using nuclear DNA, particularly using recombining genes such as *CCR*, and true monophyletic groups may not be found unless recombination is taken into consideration. It may be that several independent nuclear genes in recombination cold-spots (regions of low recombination) will be more useful for phylogenetic reconstruction. However, recombination is a major generator of sequence diversity and has most likely contributed substantially to much of the allelic variation observable in extant species. Discovering genomic regions in recombination cold-spots with sufficient variation for reconstruction of species-level phylogenies may

be problematic (Posada and Crandall, 2001; Tenaillon et al., 2001; Zhang and Hewitt, 2003).

Recombination hotspots are usually genomic regions at which meiotic crossover junctions are found at high frequency (Okagaki and Weil, 1997). Two potential recombination hotspots were found within the *CCR* gene fragment (Figure 5). The first of these (between alignment positions 230-451) coincided with the exon 4-intron 4 boundary region and the second (between alignment positions 890-1125) was within the latter half of intron 4. The occurrence of both of these breakpoint regions within the intron is interesting. In *Eucalyptus CCR*, intron 4 is the largest intron (Lacombe et al., 1997) and it has been suggested that large introns favour intragenic recombination (Duret, 2001). A breakpoint region around the exon 4-intron 4 boundary may explain the large number of sequence polymorphisms in exon 4, which are at a similar frequency to those of the intron, and are two times more frequent than in exon 5 (Table 3). The breakpoint positions occurring in the one intron may also explain the presence of the large number of indels there, a phenomenon suggested to be associated with a high recombination rate (Duret, 2001). McKinnon et al. (in press) identified two potential *CCR* recombinants in *E. globulus*, and the breakpoint for one of these was consistent with the first hotspot detailed here, while the other occurred approximately 430 base pairs into intron two. In maize, recombination junctions have been found to resolve at hotspots (Patterson et al., 1995; Xu et al., 1995), at specific sequence regions (Eggleston et al., 1995) or uniformly across a locus (Dooner and Martínez-Férez, 1997; Okagaki and Weil, 1997), with the mechanisms determining these various patterns remaining unidentified. It is unknown whether *CCR* has a uniform distribution of crossover junctions, or whether recombination in this gene has polarity (i.e. its frequency is higher at one end of the gene than the other) as is found in some maize genes (Eggleston et al., 1995; Patterson et al., 1995; Xu et al., 1995). The potential breakpoint found in intron 2 of *CCR* by McKinnon et al. (in press) suggests that there may be multiple recombination hotspots across the *CCR* gene.

Previous reports of intragenic recombination in plants have been confined to *Arabidopsis* (Innan et al., 1996), *Petunia* (Wang et al., 2001), *Zea* (Dooner and Martínez-Férez, 1997; Hulbert, 1997; Okagaki and Weil, 1997) and *Brassica*

(Kusaba et al., 1997; Awadalla and Charlesworth, 1999), and have involved recombination within a single species. These studies have mostly included genes involved in self-incompatibility (*SLG*), anthocyanin pigmentation (*bz*), starch synthesis (*waxy*), and alcohol dehydrogenase (*Adh*). The current study is one of the first reports detailing the use of a single copy nuclear gene in phylogenetic analysis of the eucalypts. It is the first detailed report of intragenic recombination in both *CCR* and *Eucalyptus*, and also is the first indication of intragenic recombination involving species that are currently classified in different sections of a plant genus. The occurrence of intragenic recombination in *CCR* not only has implications for phylogenetic reconstruction, but because the gene has an important role in lignin biosynthesis, there is possible functional significance. Recombination, through the creation of mosaic gene structures, generates allelic diversity. By effectively generating mosaic proteins (Posada and Crandall, 2001) recombination has the potential to produce functional diversity in proteins. Protein diversity may lead to variation in the lignin profiles of different eucalypt species. The results presented here may be of particular interest to the forestry industry since the detectable recombinants, *E. globulus*, *E. grandis* and *E. saligna*, are some of the main eucalypt species used for pulp production. Recombination breaks down linkage disequilibrium (Tenaillon et al., 2001) and the results of the current study are in concordance with low levels of linkage disequilibrium found in *CCR* of *E. nitens* (Thumma et al., 2005). The breakdown of linkage disequilibrium has important implications for association studies (Posada and Crandall, 2001). Low linkage disequilibrium may mean that a positive association between a mutation and variation in a trait may involve the causative mutation, and is not due to linkage to another mutation further up- or down-stream.

In conclusion, phylogenetic analysis using a single copy nuclear gene, *CCR*, was unable to resolve three sections of *Eucalyptus* into monophyletic groups, similar to previous studies using nrDNA. This is probably due to different parts of the gene having different evolutionary histories because of the occurrence of intragenic recombination, although paraphyly of section *Latoangulatae* can not be discounted. Recombination involved the three sections of eucalypts analysed here and was possibly facilitated by the naturally occurring hybridisation between these sections. Collectively the results suggest that the use of *CCR* for phylogenetic analysis of the

eucalypts will be problematic unless recombination is considered. Analysis of low copy genes not influenced by recombination may produce a better understanding of the phylogeny of these eucalypt sections.

CONCLUSIONS

This study evaluated genetic variation in the wood chemistry of *Eucalyptus globulus* and aimed to identify molecular variation that may alter lignin. Upon review of the genomic research that has been conducted in *Eucalyptus* to date, chemical wood traits were found to be a major research topic, particularly lignin. Many of the lignin biosynthesis and transcription factor genes have been cloned, characterised and positioned on genetic linkage maps. In addition, quantitative trait loci (QTL) have been identified for both lignin content and composition, some of which collocate with these mapped genes. Large EST projects have also obtained sequence for the "complete" set of 21 lignin biosynthesis genes for *Eucalyptus*. Furthermore, the creation of transgenic eucalypts down-regulated for lignin biosynthesis genes have aided in the identification of genes likely to control lignin. Collectively, these approaches have given an enormous boost to deciphering the genetic control of lignin. High levels of sequence polymorphism have also been found in two lignin biosynthesis genes in *E. globulus* (Poke et al., 2003), which has opened the way for association studies aimed to find linkages between DNA variation and lignin variation. The sequencing of the *E. camaldulensis* genome will be a valuable resource for answering questions in *E. globulus*, with the extrapolation of gene and QTL data allowing genes underlying QTL to be discovered, gene function characterised by identifying orthologs in *Arabidopsis* and other model species, and regulatory elements discovered.

Before gene markers associated with improved wood quality can be employed in eucalypt breeding programs, large numbers of trees must be screened to assess natural variation in the traits of interest. Near infrared reflectance (NIR) analysis proved to be an efficient, reliable and cost-effective tool for predicting the wood chemistry of *E. globulus*, using either ground or solid wood samples. This was the first published report of the prediction of lignin and extractives contents from ground wood in *E. globulus*, with good calibration statistics, and strong correlations between laboratory values and NIR predictions. For the first time in *Eucalyptus*, successful calibrations were also developed for the prediction of lignin, extractives and cellulose contents from solid wood. Calibration statistics were good for all traits

(with the exception of acid-soluble lignin content), and correlations between predicted and laboratory values were strong. Using NIR analysis on solid wood will further increase the rapidity of trait prediction, eliminating the need for the time-consuming wood-grinding step of the procedure. These calibrations will not only be of benefit to the forestry industry, but will allow the screening of the large sample sizes required for association studies.

Determination of within-tree variation in lignin, extractives and cellulose contents in *E. globulus* using the NIR calibrations developed for solid wood, showed that the current sampling strategy of taking increment wood cores at breast height, initially developed for pulp yield, is also effective for these traits. While breast height wood cores may not be optimal for obtaining whole tree estimates of extractives content, maximum values are likely to be determined, which may be a sufficient measure for breeding programs. Longitudinal patterns of variation were consistent with previous reports for this species, and patterns of radial variation supported those found for pulp yield in *E. globulus*. Different patterns of within-tree wood chemistry variation reported for the eucalypt species, particularly *E. globulus* and *E. nitens*, suggests there may be differences in the genetic control of wood chemistry between them. This may indicate that extrapolation of molecular results between these species may not be possible.

Genetic variation was detected in the base population trial of *E. globulus* for lignin content and decay. The significant differences found between localities, suggested improvement could be made in these traits through selection of the best localities in breeding programs. Narrow-sense heritabilities were low for all traits except acid-soluble lignin content, however high family means heritabilities suggested gain could be made through selection of the best families across all localities. The correlations identified between traits indicated that during selection for the breeding objective traits (basic density, pulp yield and volume) it is likely that favourable states in the chemical wood properties, decay resistance and fibre properties will be selected concurrently, whereas growth must be selected for independently. This indicated that the inclusion of traits such as lignin in the selection criteria of breeding programs may be unnecessary. However, the presence of genetic variation

and moderate heritability estimates suggest that breeding for improvement in lignin content is feasible.

The identification of DNA markers that could be used to screen plants for wood quality at an early age may be beneficial in tree breeding. Currently, lignin can only be measured in older trees, which is not ideal. No relationship was found between a *CCR* amino acid substitution and variation in lignin content, lignin composition or density. However, a genotype by site interaction for lignin composition may be present, which requires further exploration. Lignin content and composition are traits likely to have large environmental components, produced in response to factors such as decay and the amount of tension wood. Between site differences are therefore likely. Furthermore, it is possible that intron polymorphisms that cause splice site variants may cause significant changes to the protein, rather than missense mutations in the exons, which would be consistent with the associations found between *CCR* SNPs and variation in microfibril angle (Thumma et al., 2005). Variation in lignin profiles may also be caused by the cumulative effect of several missense mutations in the one allele. *CCR* has been found to have low levels of linkage disequilibrium in *Eucalyptus* (Chapter 7; Thumma et al., 2005), which is expected for species with high outcrossing rates. This means adjacent SNPs are linked in small regions only, and the causative variation is more likely to be identified when positive associations are found. This is in contrast to species such as *Arabidopsis thaliana* that are selfing and have high levels of linkage disequilibrium.

Not all of the SNPs that have been found in *CCR* could be explored for an association with lignin variation in this study, but the considerable levels of DNA polymorphism suggested this gene could be a good phylogenetic tool for resolving relationships among *Eucalyptus* species. The *CCR* sequence could not resolve three sections of subgenus *Symphyomyrtus*, *Latoangulatae, Exsertaria* and *Maidenaria*, into monophyletic groups. However, it did provide more resolution between the former two sections than has been found previously using nuclear encoded ribosomal DNA. This suggested that these sections may indeed be monophyletic, and if the right nuclear gene could be found in an area of low recombination, low copy number nuclear genes could be useful in resolving eucalypt phylogeny. The phylogeny obtained with *CCR* was likely confounded by recombination which may

explain the anomalous positions of some species in the phylogenetic tree. The two statistically significant intragenic recombination events detected involved species from different sections, which have probably been facilitated by intersectional hybridisation. Of the eucalypt sections, these three have the highest incidence of naturally occurring intersectional hybridisation. Ancient recombination events are likely to make elucidation of the phylogenetic relationships among *Eucalyptus* difficult. More recent recombination events are quite exciting and suggest there will be great intraspecific variation in *CCR*, as has already been shown for *E. globulus* (outlined in Chapter 7). This also suggests that large variations in lignin may be generated through recombinant proteins. This was the first detailed report of intragenic recombination in both *CCR* and *Eucalyptus*, and between species currently classified in different sections of a plant genus.

During this study, techniques for quickly and cost-effectively measuring wood chemistry in *E. globulus* were developed, which will be of great benefit to the forestry industry. These allowed a primary analysis of the natural genetic variation in the chemical wood properties of *E. globulus*, with significant differences found between localities, and heritability estimates also obtained which will provide a useful addition to the knowledge base of the species. Identification of the relationships between traits, both phenotypically and genetically will allow the potential effect of selection on other wood properties to be estimated. Collectively, the results suggested that breeding for improvement in these traits could be possible. However, the cause(s) of lignin variation at the DNA level remains elusive. As yet no DNA marker has been found which can assist tree breeders to more quickly screen for favourable lignin profiles. The genetic architecture of *CCR* seems to be quite complicated, with intragenic recombination contributing to the high allelic diversity of this gene. Therefore many SNP combinations will be possible, for which the effect on lignin may be cumulative. Recombinant alleles seem to be shared by several important forestry species, and therefore it may be that species differences in lignin may not be as large as initially thought. The full extent of intragenic recombination in *CCR* still remains to be characterised, but may eventually help elucidate the true evolutionary relationships among the *Eucalyptus* species.

REFERENCES

Agrama H.A., George T.L., Salah S.F. (2002) Construction of a genome map for *Eucalyptus camaldulensis* DEHN. Silvae Genetica 51(5-6): 201-206.

Ahuja M.R. (2001) Recent advances in molecular genetics of forest trees. Euphytica 121(2): 173-195.

Alves E., Velini E.D., Trindade M.L.B., Cataneo A.C., Marino C.L., Mori E.S., Guerrini I.A., Furtado E.L., Wilcken C.F. (2005) *Eucalyptus* ESTs related to genes for oxidative stress. Journal of Environmental Science and Health, Part B: Pesticides, Food Contaminants, and Agricultural Wastes 40(1): 151-157.

Apiolaza L.A., Raymond C.A., Yeo B.J. (in press) Genetic variation of physical and chemical wood properties of *Eucalyptus globulus*. Silvae Genetica 54(4-5): 160-166.

Aritome Y., Baba K.I., Itoh T., Hibino T. (1999) cDNA analysis of the genes increasing their expression on tension wood formation in *Eucalyptus camaldulensis* L. Wood Research 86: 39-40.

Appita Standard (1978) Klason lignin in wood and pulp. Australian Pulp and Paper Industry Technical Association, Appita Standard P11s.

Appita Standard (1994) Organic solvent extractives in wood, pulp and paper. Australian Pulp and Paper Industry Technical Association, Appita Standard 1301.012s.

Awadalla P., Charlesworth D. (1999) Recombination and selection at Brassica self-incompatibility loci. Genetics 152(1): 413-425.

Axrup L., Markides K., Nilsson T. (2000) Using miniature diode array NIR spectrometers for analysing wood chips and bark samples in motion. Journal of Chemometrics 14: 561-572.

Baillères H., Davrieus F., Pichavant F.H. (2002) Near infrared analysis as a tool for rapid screening of some major wood characteristics in a eucalyptus breeding program. Annals of Forest Science 59(5-6): 479-490.

Barbour R.C., Potts B.M., Vaillancourt R.E. (2003) Gene flow between introduced and native *Eucalyptus* species: exotic hybrids are establishing in the wild. Australian Journal of Botany 51(4): 429-439.

Barros E., Verryn S., Hettasch M. (2002) Identification of PCR-based markers linked to wood splitting in *Eucalyptus grandis*. Annals of Forest Science 59(5-6): 675-678.

Barry K.M., Pearce R.B., Evans S.D., Hall L.D., Mohammed C.M. (2001) Initial defence responses in sapwood of *Eucalyptus nitens* (Maiden) following wounding and fungal innoculation. Physiological and Molecular Plant Pathology 58(2): 63-72.

Barry K.M., Pearce R.B., Mohammed C.L. (2000) Properties of reaction zones associated with decay from pruning wounds in plantation-grown *Eucalyptus nitens*. Forest Pathology 30: 233-245.

Baucher M., Halpin C., Petit-Conil M., Boerjan W. (2003) Lignin: Genetic engineering and impact on pulping. Critical Reviews in Biochemistry and Molecular Biology 38(4): 305-350.

Baucher M., Monties B., Van Montagu M., Boerjan W. (1998) Biosynthesis and genetic engineering of lignin. Critical Reviews in Plant Sciences 17(2): 125-197.

Beavis W.D. (1998) QTL analyses: power, precision, and accuracy. *In* H.A. Patterson, ed, Molecular dissection of complex traits. CRC Press, Boca Raton, Fla, pp 145-162.

Bhoora R., Berger D.K., Wingfield M.J., Wingfield B.D. (2003) PCR cloning by genome walking of a complete polygalacturonase-inhibiting protein gene from *Eucalyptus grandis*. South African Journal of Science 99(9-10): 410-412.

Boerjan W., Ralph J., Baucher M. (2003) Lignin biosynthesis. Annual Review of Plant Biology 54: 519-546.

Bogunic F., Muratovic E., Brown S.C., Siljak-Yakovlev S. (2003) Genome size and base composition of five *Pinus* species from the Balkan region. Plant Cell Reports 22(1): 59-63.

Boiffin V., Hodges M., Galvez S., Balestrini R., Bonfante P., Gadal P., Martin F. (1998) Eucalypt NADP-dependent isocitrate dehydrogenase - cDNA cloning and expression in ectomycorrhizae. Plant Physiology 117(3): 939-948.

Bootle K.R. (1983) Wood in Australia: types, properties, uses. McGraw-Hill, Sydney.

Bossinger G., Leitch M. (2000) Isolation of cambium-specific genes from *Eucalyptus globulus* Labill. *In* R. Savidge, J. Barnett, R. Napier, eds, Cell and molecular biology of wood formation. BIOS Scientific Ltd, Oxford, UK, pp 203-207.

Bossinger G., Leitch M.A. (1999) *HEC1*, a Knotted1-type homeobox gene from eucalypts and the molecular control of cambial activity. Journal of Experimental Botany 50: 23-24.

Boudet A.M., Grima-Pettenati J., Feuillet C., Goffner D. (1993) Cinnamyl alcohol dehydrogenase from *Eucalyptus* - molecular cloning, expression and gene characterization. Journal of Cellular Biochemistry Suppl: 26.

Bradshaw H.D., Grattapaglia D. (1994) QTL mapping in interspecific hybrids of forest trees. Forest Genetics 1(4): 191-196.

Brill E.M., Watson J.M. (2004) Ectopic expression of a *Eucalyptus grandis SVP* orthologue alters the flowering time of *Arabidopsis thaliana*. Functional Plant Biology 31(3): 217-224.

Brondani R.P.V., Brondani C., Grattapaglia D. (2002) Towards a genus-wide reference linkage map for *Eucalyptus* based exclusively on highly informative microsatellite markers. Molecular Genetics and Genomics 267(3): 338-347.

Brondani R.P.V., Brondani C., Tarchini R., Grattapaglia D. (1998) Development, characterization and mapping of microsatellite markers in *Eucalyptus grandis* and *E. urophylla*. Theoretical and Applied Genetics 97(5-6): 816-827.

Brondani R.P.V., Grattapaglia D. (1996a) Comparative mapping in *Eucalyptus* reveals significant RAPD locus linkage and order conservation between trees and greater meiotic recombination in females. *In* Plant Genome IV Conference. San Diego, CA, Jan 14-18.

Brondani R.P.V., Grattapaglia D. (1996b) Mapped RAPD markers are transferable among *Eucalyptus* trees from the same population. *In* Plant Genome IV Conference. San Diego, CA, Jan 14-18.

Brooker M.I.H. (2000) A new classification of the genus *Eucalyptus* L'Her. (*Myrtaceae*). Australian Systematic Botany 13: 79-148.

Brown G.R., Bassoni D.L., Gill G.P., Fontana J.R., Wheeler N.C., Megraw R.A., Davis M.F., Sewell M.M., Tuskan G.A., Neale D.B. (2003) Identification of

quantitative trait loci influencing wood property traits in loblolly pine (*Pinus taeda* L.). III. QTL verification and candidate gene mapping. Genetics 164(4): 1537-1546.

Brummitt R.K. (2002) How to chop up a tree. Taxon 51: 31-41.

Brunner A.M., Busov V.B., Strauss S.H. (2004) Poplar genome sequence: functional genomics in an ecologically dominant plant species. Trends in Plant Science 9(1): 49-56.

Bundock P.C. (2003) Genome mapping in *Eucalyptus globulus*. PhD Thesis. University of Tasmania, Hobart, Australia.

Bundock P.C., Hayden M., Vaillancourt R.E. (2000) Linkage maps of *Eucalyptus globulus* using RAPD and microsatellite markers. Silvae Genetica 49(4-5): 223-232.

Byrne M., Hines B. (2004) Phylogeographical analysis of cpDNA variation in *Eucalyptus loxophleba* (*Myrtaceae*). Australian Journal of Botany 52(4): 459-470.

Byrne M., Macdonald B. (2000) Phylogeography and conservation of three oil mallee taxa, *Eucalyptus kochii* ssp *kochii*, ssp *plenissima* and *E. horistes*. Australian Journal of Botany 48(3): 305-312.

Byrne M., Marquez-Garcia M.I., Uren T., Smith D.S., Moran G.F. (1996) Conservation and genetic diversity of microsatellite loci in the genus *Eucalyptus*. Australian Journal of Botany 44: 331-341.

Byrne M., Moran G.F. (1994) Population divergence in the chloroplast genome of *Eucalyptus nitens*. Heredity 73: 18-28.

Byrne M., Moran G.F., Tibbits W.N. (1993) Restriction map and maternal inheritance of chloroplast DNA in *Eucalyptus nitens*. Journal of Heredity 84(3): 218-220.

Byrne M., Murrell J.C., Allen B., Moran G.F. (1995) An integrated genetic-linkage map for eucalypts using RFLP, RAPD and isozyme markers. Theoretical and Applied Genetics 91(6-7): 869-875.

Byrne M., Murrell J.C., Owen J.V., Kriedemann P., Williams E.R., Moran G.F. (1997a) Identification and mode of action of quantitative trait loci affecting seedling height and leaf area in *Eucalyptus nitens*. Theoretical and Applied Genetics 94(5): 674-681.

Byrne M., Murrell J.C., Owen J.V., Williams E.R., Moran G.F. (1997b) Mapping of quantitative trait loci influencing frost tolerance in *Eucalyptus nitens*. Theoretical and Applied Genetics 95(5-6): 975-979.

Campinhos E.N., Bertolucci F.L., Alfenas A.C., Grattapaglia D. (1996) Stability patterns of height growth QTLs across variable genetic backgrounds in *Eucalyptus grandis*. *In* Plant Genome IV Conference. San Diego, CA, Jan 14-18.

Cardon L.R., Bell J.I. (2001) Association study designs for complex diseases. Nature Reviews Genetics 2(2): 91-99.

Chabannes M., Barakate A., Lapierre C., Marita J.M., Ralph J., Pean M., Danoun S., Halpin C., Grima-Pettenati J., Boudet A.M. (2001) Strong decrease in lignin content without significant alteration of plant development is induced by simultaneous down-regulation of cinnamoyl CoA reductase (CCR) and cinnamyl alcohol dehydrogenase (CAD) in tobacco plants. The Plant Journal 28(3): 257-270.

Chaparro J.X., Mitchelson K.R., Russell H., Gibbings M., Smail G., Li Y., Teasdale R. (1997) Genetic mapping and QTL detection in the tropical species *Eucalyptus grandis* and *E. urophylla* and their hybrids. *In* Plant & Animal Genome V Conference. San Diego, CA, Jan 12-16.

Chappill J.A., Ladiges P.Y. (1996) Phylogenetic analysis of *Eucalyptus* informal subgenus *Symphyomyrtus* section *Maidenaria*. Australian Systematic Botany 9(1): 71-93.

Chen D.M., De Filippis L.F. (2001) Differentially expressed genes identified during salt adaptation in *Eucalyptus microcorys*: down-regulation of a cDNA sequence coding for alpha-tubulin. Journal of Plant Physiology 158(9): 1195-1202.

Chen Z., Chang S., Ho C., Chen Y., Tsai J., Chiang V.L. (2001) Plant production of transgenic *Eucalyptus camaldulensis* carrying the *Populus tremuloides* cinnamate 4-hydroxylase gene. Taiwan Journal of Forest Science 16(4): 249-258.

Chimwamurombe P.M., Botha A.M., Wingfield M.J., Wingfield B.D. (2001) Molecular relatedness of the polygalacturonase-inhibiting protein genes in *Eucalyptus* species. Theoretical and Applied Genetics 102(4): 645-650.

Cotterill P.P., Brolin A. (1997) Improving *Eucalyptus* wood, pulp and paper quality by genetic selection. *In* Proceedings of the IUFRO Conference on Silviculture and Improvement of Eucalypts, pp 1-13. Salvador, Brazil, Aug 24-29.

Dale G.T., Aitken K.S., Sasse J.M. (2000) Development of salt-tolerant *E. camaldulensis* x *E. grandis* hybrid clones using phenotypic selection and genetic mapping. *In* Proceedings of QFRI/CRC-SPF Symposium on Hybrid Breeding and Genetics of Forest Trees, pp 227-233. Noosa, QLD, Apr 9-14.

Danilevskaya O.N., Hermon P., Hantke S., Muszynski M.G., Kollipara K., Ananiev E.V. (2003) Duplicated *fie* genes in maize: Expression pattern and imprinting suggest distinct functions. Plant Cell 15(2): 425-438.

De Melis L.E., Whiteman P.H., Stevenson T.W. (1999) Isolation and characterisation of a cDNA clone encoding cinnamyl alcohol dehydrogenase in *Eucalyptus globulus* Labill. Plant Science 143(2): 173-182.

Decroocq V., Zhu X.M., Kauffman M., Kyozuka J., Peacock W.J., Dennis E.S., Llewellyn D.J. (1999) A TM3-like MADS-box gene from *Eucalyptus* expressed in both vegetative and reproductive tissues. Gene 228(1-2): 155-160.

Diaz E.C., Martin F., Tagu D. (1996) Eucalypt alpha-tubulin: cDNA cloning and increased level of transcripts in ectomycorrhizal root system. Plant Molecular Biology 31(4): 905-910.

Domingues D.S., Leite S.M.M., Coscrato V.E., Farro A.P.C., Maia I.G., Velini E.D., Furtado E.L., Wilcken C.F., Guerrini I.A., Mori E.S., Camargo L.E.A., Marino C.L. (2004) ESTs related to phloem retranslocation of boron in *Eucalyptus*. *In* Plant & Animal Genomes XII Conference. San Diego, CA, Jan 10-14.

Dooner H.K., Martínez-Férez I.M. (1997) Recombination occurs uniformly within the bronze gene, a meiotic recombination hotspot in the maize genome. Plant Cell 9(9): 1633-1646.

Doughty R.W. (2000) The *Eucalyptus*: A natural and commercial history of the gum tree. John Hopkins University Press, Baltimore, MD.

Doyle J.J., Doyle J.L. (1990) Isolation of plant DNA from fresh tissue. Focus 12: 13-15.

Draghici S., Khatri P., Shah A., Tainsky M.A. (2003) Assessing the functional bias of commercial microarrays using the onto-compare database. Biotechniques Suppl: 55-61.

Duggan D.J., Bittner M., Chen Y.D., Meltzer P., Trent J.M. (1999) Expression profiling using cDNA microarrays. Nature Genetics 21(Suppl. S): 10-14.

Duplessis S., Courty P.E., Tagu D., Martin F. (2005) Transcript patterns associated with ectomycorrhiza development in *Eucalyptus globulus* and *Pisolithus microcarpus*. New Phytologist 165(2): 599-611.

Duret L. (2001) Why do genes have introns? Recombination might add a new piece to the puzzle. Trends in Genetics 17(4): 172-175.

Dutkowski G.W., Gilmour A.R., Borralho N.M.G. (2001) Modification of the additive relationship matrix for open pollinated trials. *In* IUFRO International Symposium on Developing the Eucalypt of the Furture. Valdivia, Chile, Sep 10-15.

Dutkowski G.W., Potts B.M. (1999) Geographic patterns of genetic variation in *Eucalyptus globulus* ssp. *globulus* and a revised racial classification. Australian Journal of Botany 47(2): 237-263.

Eggleston W.B., Alleman M., Kermicle J.L. (1995) Molecular organization and germinal instability of *R-stippled* maize. Genetics 141: 347-360.

Eisen M.B., Spellman P.T., Brown P.O., Botstein D. (1998) Cluster analysis and display of genome-wide expression patterns. Proceedings of the National Academy of Sciences USA 95(25): 14863-14868.

Eldridge K., Davidson J., Harwood C., van Wyk G. (1993) Eucalypt domestication and breeding. Clarendon Press, Oxford, UK.

Emshwiller E., Doyle J.J. (1999) Chloroplast-expressed glutamine synthetase (ncpGS): Potential utility for phylogenetic studies with an example from *Oxalis* (Oxalidaceae). Molecular Phylogenetics and Evolution 12(3): 310-319.

Evans R., Hughes M., Menz D. (1999) Microfibril angle variation by scanning X-ray diffractometry. Appita Journal 52(5): 363-367.

Evans R., Stringer S., Kibblewhite R.P. (2000) Variation of microfibril angle, density and fibre orientation in twenty-nine *Eucalyptus nitens* trees. Appita Journal 53(5): 450-457.

Fairbairn D.J., Liu W.H., Schachtman D.P., Gomez-Gallego S., Day S.R., Teasdale R.D. (2000) Characterisation of two distinct HKT1-like potassium transporters from *Eucalyptus camaldulensis*. Plant Molecular Biology 43(4): 515-525.

Faridah E. (2004) Molecular and physiological studies of salt tolerance in eucalypts. PhD Thesis. University of Melbourne, Melbourne, Australia.

Felsenstein J. (1989) PHYLIP - Phylogeny inference package (Version 3.2). Cladistics 5: 164-166.

Feuillet C., Boudet A.M., Grima-Pettenati J. (1993) Nucleotide sequence of a cDNA encoding cinnamyl alcohol dehydrogenase from *Eucalyptus*. Plant Physiology 103(4): 1447-1447.

Flint-Garcia S.A., Thornsberry J.M., Buckler E.S. (2003) Structure of linkage disequilibrium in plants. Annual Review of Plant Biology 54: 357-374.

Freeman J.S., Jackson H.D., Steane D.A., McKinnon G.E., Dutkowski G.W., Potts B.M., Vaillancourt R.E. (2001) Chloroplast DNA phylogeography of *Eucalyptus globulus*. Australian Journal of Botany 49(5): 585-596.

Freeman J.S., Milgate A.W., Vaillancourt R.E., Mohammed C.L., Potts B.M. (2003) Genetic control of resistance to *Mycosphaerella* leaf disease in *Eucalyptus globulus*. *In* 19th Genetics Congress, p 155. Melbourne, Australia, Jul 6-11.

Fullard K.J., Moran G. (2003) Identification of frost tolerance genes in a *Eucalyptus* hybrid cross. *In* IUFRO Tree Biotechnology Meeting. Umeå, Sweden, Jun 7-12.

Furtado E.L., Rosa D.D., Zamprogno K., Marino C.L., Mori E.S., Guerrini I.A., Wilcken C.F., Maia I.G., Velini E., Camargo L.E.A. (2004) Pathogenesis related proteins in *Eucalyptus* genome. *In* Plant & Animal Genomes XII Conference. San Diego, CA, Jan 10-14.

Gan S.M., Shi J.S., Li M., Wu K.M., Wu J.Y., Bai J.Y. (2003) Moderate-density molecular maps of *Eucalyptus urophylla* S. T. Blake and *E. tereticornis* Smith genomes based on RAPD markers. Genetica 118(1): 59-67.

Gardiner C.A., Crawford D.F. (1987) Seed collections of *Eucalyptus globulus* subsp. *globulus* for tree improvement purposes. CSIRO Division of Forest Research, Canberra.

Gardiner C.A., Crawford D.F. (1988) Seed collections of *Eucalyptus globulus* subsp. *globulus* Labill. for tree improvement purposes. CSIRO Division of Forestry and Forest Products, Canberra.

Gilmour A.R., Cullis B.R., Welham S.J., Thompson R. (1999) ASREML computer software. NSW Agriculture, Orange.

Gindl W., Teischinger A., Schwanninger M., Hinterstoisser B. (2001) The relationship between near infrared spectra of radial wood surfaces and wood mechanical properties. Journal of Near Infrared Spectroscopy 9(4): 255-261.

Gion J.M. (2000) Mapping eligible genes in *Eucalyptus*. Bois et Forets des Tropiques 266: 61-65.

Gion J.M., Boudet C., Grima-Pettenati J., Pichavant F.H., Plomion C., Bailleres H., Verhaegen D. (2001) A candidate genes approach identifies *CCR*, *PAL* and *C4H* as loci for syringyl/guaiacyl ratio in a interspecific hybrid between *E. urophylla* and *E. grandis*. *In* IUFRO International Symposium on Developing the Eucalypt of the Future. Valdivia, Chile, Sep 10-15.

Gion J.M., Rech P., Grima-Pettenati J., Verhaegen D., Plomion C. (2000) Mapping candidate genes in *Eucalyptus* with emphasis on lignification genes. Molecular Breeding 6(5): 441-449.

Glaubitz J.C., Emebiri L.C., Moran G.F. (2001) Dinucleotide microsatellites from *Eucalyptus sieberi*: inheritance, diversity, and improved scoring of single-base differences. Genome 44(6): 1041-1045.

Goff S.A., Ricke D., Lan T.H., Presting G., Wang R.L., Dunn M., Glazebrook J., Sessions A., Oeller P., Varma H., Hadley D., Hutchinson D., Martin C., Katagiri F., Lange B.M., Moughamer T., Xia Y., Budworth P., Zhong J.P., Miguel T., Paszkowski U., Zhang S.P., Colbert M., Sun W.L., Chen L.L., Cooper B., Park S., Wood T.C., Mao L., Quail P., Wing R., Dean R., Yu Y.S., Zharkikh A., Shen R., Sahasrabudhe S., Thomas A., Cannings R., Gutin A., Pruss D., Reid J., Tavtigian S., Mitchell J., Eldredge G., Scholl T., Miller R.M., Bhatnagar S., Adey N., Rubano T., Tusneem N., Robinson R., Feldhaus J., Macalma T., Oliphant A., Briggs S. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp *japonica*). Science 296(5565): 92-100.

Goffner D., Grima-Pettenati J., Feuillet C., Boudet A.M. (1995) Cinnamoyl-CoA reductase and cinnamyl alcohol dehydrogenase: characterization at the biochemical and molecular level of two enzymes involved in lignin

synthesis. *In* in Polyphenols 94: 17th International Conference, Palma de Mallorca, Spain, 23-27 May, 1994. Institut National de la Recherche Agronomique, Paris, France

Goffner D., Van Doorsselaere J., Yahiaoui N., Samaj J., Grima-Pettenati J., Boudet A.M. (1998) A novel aromatic alcohol dehydrogenase in higher plants: molecular cloning and expression. Plant Molecular Biology 36(5): 755-765.

Gominho J., Rodrigues J., Almeida M.H., Leal A., Cotterill P.P., Pereira H. (1997) Assessment of pulp yield and lignin content in a first-generation clonal testing of *Eucalyptus globulus* in Portugal. *In* Proceedings of the IUFRO Conference on Silviculture and Improvement of Eucalypts, pp 84-89. Salvador, Brazil, Aug 24-29.

Gonzalez E.R., de Andrade A., Bertolo A.L., Lacerda G.C., Carneiro R.T., Defavari V.A.P., Labate M.T.V., Labate C.A. (2002) Production of transgenic *Eucalyptus grandis* x *E. urophylla* using the sonication-assisted *Agrobacterium* transformation (SAAT) system. Functional Plant Biology 29(1): 97-102.

Goujon T., Ferret V., Mila I., Pollet B., Ruel K., Burlat V., Joseleau J.P., Barriere Y., Lapierre C., Jouanin L. (2003) Down-regulation of the *AtCCR1* gene in *Arabidopsis thaliana*: effects of phenotype, lignins and cell wall degradability. Planta 217: 218-228.

Grattapaglia D. (2002) Integrating genomic biotechnologies into genetic improvement of *Eucalyptus*. *In* Universidad de Vigo, ed, Socioeconomia, patologia, tecnologia y sostenibilidade del eucalipto, Vol 1. Graficas Diumaro, pp 101-122.

Grattapaglia D. (2003) *Eucalyptus globulus*: key resource for *Eucalyptus* genomics research. *In* Proceedings of the IUFRO Symposium "Simposio Iberoamericano de *Eucalyptus globulus*". Montevideo, Uruguay, Oct 30-31.

Grattapaglia D., Bertolucci F.L., Sederoff R.R. (1995) Genetic-mapping of QTLs controlling vegetative propagation in *Eucalyptus grandis* and *E. urophylla* using a pseudo-testcross strategy and RAPD markers. Theoretical and Applied Genetics 90(7-8): 933-947.

Grattapaglia D., Bertolucci F.L.G., Penchel R., Sederoff R.R. (1996) Genetic mapping of quantitative trait loci controlling growth and wood quality traits

in *Eucalyptus grandis* using a maternal half-sib family and RAPD markers. Genetics 144(3): 1205-1214.

Grattapaglia D., Bradshaw H.D. (1994) Nuclear DNA content of commercially important *Eucalyptus* species and hybrids. Canadian Journal of Forest Research 24: 1074-1078.

Grattapaglia D., Sederoff R. (1994) Genetic-linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross mapping strategy and RAPD markers. Genetics 137(4): 1121-1137.

Greaves B.L., Borralho N.M.G. (1996) The influence of basic density and pulp yield on the cost of eucalypt kraft pulping: a theoretical model for tree breeding. Appita Journal 49(2): 90-93.

Greaves B.L., Borralho N.M.G., Raymond C.A. (1997) Breeding objective for plantation eucalypts grown for production of kraft pulp. Forest Science 43(4): 465-472.

Greaves B.L., Borralho N.M.G., Raymond C.A., Evans R., Whiteman P.H. (1997) Age-age correlations in, and relationships between basic density and growth in *Eucalyptus nitens*. Silvae Genetica 46(5): 264-270.

Griffin A.R., Burgess I.P., Wolf L. (1988) Patterns of natural and manipulated hybridization in the genus *Eucalyptus* L'Herit.: a review. Australian Journal of Botany 36: 41-66.

Griffiths A.J.F., Wessler S.R., Lewontin R.C., Gelbart W.M., Susuki D.T., Miller J.H. (2005) An introduction to genetic analysis, 8th Edition. W.H. Freeman, New York.

Grima-Pettenati J., Feuillet C., Goffner D., Borderies G., Boudet A.M. (1993) Molecular cloning and expression of a *Eucalyptus gunnii* cDNA clone encoding cinnamyl alcohol dehydrogenase. Plant Molecular Biology 21(6): 1085-1095.

Grima-Pettenati J., Goffner D. (1999) Lignin genetic engineering revisited. Plant Science 145(2): 51-65.

Guerrini I.A., Trigueiro R.M., Leite R.M., Mori E.S., Velini E.D., Furtado E.L., Wilcken C.F., Marino C.L., Maia I.G., Camargo L.E.A. (2004) Data mining of eucalypt ESTs related with 9-cis-epoxicarotenoide dioxigenase of ABA biosynthesis. *In* Plant & Animal Genomes XII Conference. San Diego, CA, Jan 10-14.

Harcourt R.L., Kyozuka J., Floyd R.B., Bateman K.S., Tanaka H., Decroocq V., Llewellyn D.J., Zhu X., Peacock W.J., Dennis E.S. (2000) Insect- and herbicide-resistant transgenic eucalypts. Molecular Breeding 6(3): 307-315.

Hatey F., Tosser-Klopp G., Clouscard-Martinato C., Mulsant P., Gasser F. (1998) Expressed sequence tags for genes: a review. Genetics, Selection, Evolution 30(6): 521-541.

Hawkins S., Boudet A. (2003) 'Defence lignin' and hydroxycinnamyl alcohol dehydrogenase activities in wounded *Eucalyptus gunnii*. Forest Pathology 33(2): 91-104.

Hawkins S., Boudet A., Grima-Pettenati J. (2003) Characterisation of caffeic acid O-methyltransferase and cinnamyl alcohol dehydrogenase gene expression patterns by in situ hybridisation in *Eucalyptus gunnii* Hook. plantlets. Plant Science 164(2): 165-173.

Hawkins S., Goffner D., Boudet A.M. (1994) Cinnamyl alcohol dehydrogenase polymorphism and its potential role in the control of lignin heterogeneity. Acta Horticulturae 381: 280-286.

Hawkins S.W., Boudet A.M. (1994) Purification and characterization of cinnamyl alcohol dehydrogenase isoforms from the periderm of *Eucalyptus gunnii* Hook. Plant Physiology 104(1): 75-84.

Hayden M. (1995) Mapping QTL's for growth, flowering and wood properties in *Eucalyptus globulus*. Honours Thesis. University of Tasmania, Hobart, Australia.

Hibino T., Chen J.Q., Shibata D., Higuchi T. (1994) Nucleotide sequence of a *Eucalyptus botryoides* gene encoding cinnamyl alcohol dehydrogenase. Plant Physiology 104(1): 305-306.

Hibino T., Koshiyama J., Kondo K., Doi K., Ishige N., Asada T., Sukeno S., Yamaguchi-Shinozaki K. (2002) Evaluation of transgenic eucalypts with improved *DREB1A* genes and study on the stress resistant mechanism in *Eucalyptus camaldulensis*. *In* Plant, Animal & Microbe Genomes X Conference. San Diego, CA, Jan 12-16.

Hillis W.E. (1984) Wood quality and utilization. *In* W.E. Hillis, A.G. Brown, eds, Eucalypts for wood production. Academic Press, North Ryde, NSW

Ho C., Chen Y., Chen Z. (2002) cDNA cloning and sequence analysis of 5-hydroxyconiferaldehyde O-methyltransferase from *Eucalyptus camaldulensis*. Taiwan Journal of Forest Science 17(4): 429-438.

Ho C.K., Chang S.H., Tsay J.Y., Tsai C.J., Chiang V.L., Chen Z.Z. (1998) *Agrobacterium tumefaciens*-mediated transformation of *Eucalyptus camaldulensis* and production of transgenic plants. Plant Cell Reports 17(9): 675-680.

Hodge G.R., Woodbridge W.C. (2004) Use of near infrared spectroscopy to predict lignin content in tropical and sub-tropical pines. Journal of Near Infrared Spectroscopy 12: 381-390.

Hulbert S.H. (1997) Structure and evolution of the rp1 complex conferring rust resistance in maize. Annual Review of Phytopathology 35: 293-310.

Innan H., Tajima F., Terauchi R., Miyashita N.T. (1996) Intragenic recombination in the Adh locus of the wild plant *Arabidopsis thaliana*. Genetics 143(4): 1761-1770.

Ishige N., Kondo K., Furujyo A., Hibino T. (2004) Genetic improvement for environmental stress resistance in *Eucalyptus*. *In* Plant & Animal Genomes XII Conference. San Diego, CA, Jan 10-14.

Jackson H.D., Steane D.A., Potts B.M., Vaillancourt R.E. (1999) Chloroplast DNA evidence for reticulate evolution in *Eucalyptus* (Myrtaceae). Molecular Ecology 8(5): 739-751.

Johnson G.R., Burdon R.D. (1990) Family-site interaction in *Pinus radiata*: Implications for progeny testing strategy and regionalised breeding in New Zealand. Silvae Genetica 39(2): 55-62.

Jones L., Ennos A.R., Turner S.R. (2001) Cloning and characterization of irregular xylem4 (*irx4*): a severely lignin-deficient mutant of *Arabidopsis*. The Plant Journal 26(2): 205-216.

Jones M.E., Stokoe R.L., Cross M.J., Scott L.J., Maguire T.L., Shepherd M. (2001) Isolation of microsatellite loci from spotted gum (*Corymbia variegata*), and cross-species amplification in *Corymbia* and *Eucalyptus*. Molecular Ecology Notes 1(4): 276-278.

Junghans D.T., Alfenas A.C., Brommonschenkel S.H., Oda S., Mello E.J., Grattapaglia D. (2003) Resistance to rust (*Puccinia psidii* Winter) in

Eucalyptus: mode of inheritance and mapping of a major gene with RAPD markers. Theoretical and Applied Genetics 108(1): 175-180.

Jun-ichiro T., Kei'ichi B., Takashi H., Takao I. (1997) Molecular cloning of cDNAs of the genes expressed in differentiating xylem of tension wood formation in *Eucalyptus camaldulensis* L. Wood Research 84: 31-33.

Kawaoka A., Nanto K., Sugita K., Endo S., Yamada-Watanabe K., Matsunaga E., Ebinuma H. (2003) Production and analysis of lignin-modified transgenic *Eucalyptus*. *In* IUFRO Tree Biotechnology Meeting. Umeå, Sweden, Jun 7-12.

Keane P.J., Kile G.A., Podger F.D., Brown B.N. (2000) Diseases and pathogens of eucalypts. CSIRO, Melbourne, Australia.

Keller G., Ladouce N., Teulières C., Marque C. (2004) Expression profiles of 3000 *Eucalyptus* cDNAs under cold, osmotic or salt stress. *In* Proceedings of IUFRO Conference on *Eucalyptus* in a Changing World, edited by N.M.G. Borralho, J.S. Pereira, C. Marques, J. Coutinho, M. Madeira, M. Tome, p 411. Aveiro, Portugal, Oct 11-15.

Kelley S.S., Rials T.G., Snell R., Groom L.H., Sluiter A. (2004) Use of near infrared spectroscopy to measure the chemical and mechanical properties of solid wood. Wood Science and Technology 38(4): 257-276.

Kibblewhite R.P., Evans R., Riddell M.J.C., Shelbourne C.J.A. (2004) Changes in density and wood-fibre properties with height position in 15/16-year-old *Eucalyptus nitens* and *E. fastigata*. Appita Journal 57(3): 240-247.

Kibblewhite R.P., Riddell M.J.C. (2000) Wood and kraft fibre property variation within and among nine trees of *Eucalyptus nitens*. Appita Journal 53(3): 237-244.

Kibblewhite R.P., Riddell M.J.C. (2001) Within-tree variation of some wood and kraft fibre properties of *Eucalyptus fastigata* and *E. nitens*. Appita Journal 54(2): 136-143.

Kirst M. (2003) Transcription regulation and plant diversity. PhD Thesis. North Carolina State University, Raleigh, NC.

Kirst M., Carocha V., Marques C.M., van Zyl L., Myburg A., Sederoff R. (2002) Individual genotypes of eucalypts show striking variation in xylem gene expression. *In* Plant, Animal & Microbe Genomes X Conference. San Diego, CA, Jan 12-16.

Kirst M., Myburg A.A., De Leon J.P.G., Kirst M.E., Scott J., Sederoff R. (2004) Coordinated genetic regulation of growth and lignin revealed by quantitative trait locus analysis of cDNA microarray data in an interspecific backcross of *Eucalyptus*. Plant Physiology 135(4): 2368-2378.

Kondo K., Furujyo A., Ishige N., Kasuga M., Shinozaki K., Yamaguchi-Shinozaki K., Hibino T. (2003) Analysis of the stress response genes in *Eucalyptus* and effect of introducing several stress tolerance-giving genes into *Eucalyptus* - A development situation and a practical possibility of an environmental stress resistant tree. *In* Plant & Animal Genomes XI Conference. San Diego, CA, Jan 11-15.

Kondo K., Koshiyama J., Doi K., Ishige N., Sukeno S., Hibino T., Kasuga M., Shinozaki K. (2002) Analysis of the stress response genes in *Eucalyptus* and effect of introducing a stress tolerance-giving gene (*rd29A-DREB1A*) into *Eucalyptus*. Plant and Cell Physiology 43: S113.

Kube P.D. (2005) Genetic improvement of the wood properties of *Eucalyptus nitens*: Breeding to improve solid wood and pulp properties. PhD Thesis. University of Tasmania, Hobart, Australia.

Kube P.D., Raymond C.A. (2001) Genetic parameters for *Eucalyptus nitens* solid wood traits and relationships with pulpwood traits. *In* IUFRO International Symposium on Developing the Eucalypt of the Future. Valdivia, Chile, Sep 10-15.

Kube P.D., Raymond C.A. (2002) Prediction of whole-tree basic density and pulp yield using wood core samples in *Eucalyptus nitens*. Appita Journal 55(1): 43-48.

Kusaba M., Nishio T., Satta Y., Hinata K., Ockendon D. (1997) Striking sequence similarity in inter- and intra-specific comparisons of class I SLG alleles from *Brassica oleracea* and *Brassica campestris*: Implications for the evolution and recognition mechanism. Proceedings of the National Academy of Sciences USA 94(14): 7673-7678.

Kyozuka J., Harcourt R., Peacock W.J., Dennis E.S. (1997) *Eucalyptus* has functional equivalents of the *Arabidopsis AP1* gene. Plant Molecular Biology 35(5): 573-584.

Lacombe E., Hawkins S., Van Doorsselaere J., Piquemal J., Goffner D., Poeydomenge O., Boudet A.M., Grima-Pettenati J. (1997) Cinnamoyl CoA

reductase, the first committed enzyme of the lignin branch biosynthetic pathway: Cloning, expression and phylogenetic relationships. The Plant Journal 11(3): 429-441.

Lacombe E., Van Doorsselaere J., Boerjan W., Boudet A.M., Grima-Pettenati J. (2000) Characterization of cis-elements required for vascular expression of the Cinnamoyl CoA Reductase gene and for protein-DNA complex formation. The Plant Journal 23(5): 663-676.

Lauvergeat V., Rech P., Jauneau A., Guez C., Coutos-Thevenot P., Grima-Pettenati J. (2002) The vascular expression pattern directed by the *Eucalyptus gunnii* cinnamyl alcohol dehydrogenase *EgCAD2* promoter is conserved among woody and herbaceous plant species. Plant Molecular Biology 50(3): 497-509.

Lewis C.E., Doyle J.J. (2001) Phylogenetic utility of the nuclear gene malate synthase in the palm family (*Arecaceae*). Molecular Phylogenetics and Evolution 19(3): 409-420.

Li Y., Botero J., Duong H., Clarke B., Zwart R., Teasdale R. (1999) Genetic mapping of salt tolerant eucalypt hybrids. *In* Plant & Animal Genome VII Conference. San Diego, CA, Jan 17-21.

Lincoln S.E., Daly M.J., Lander E.S. (1993) Constructing genetic linkage maps with MAPMAKER/EXP Version 3.0. *In* Whitehead Institute for Biomedical Research Technical Report, 3rd edition. Whitehead Institute, Cambridge, MA, USA

Llewellyn D.J. (2000) Herbicide tolerant forest trees. *In* S.M. Jain, S.C. Minocha, eds, Molecular biology of woody plants, Vol 2. Kluwer Academic Publishers, Dordrecht, Netherlands, pp 439-466.

Lopez G.A., Potts B.M., Dutkowski G.W., Apiolaza L.A., Gelid P.E. (2002) Genetic variation and inter-trait correlations in *Eucalyptus globulus* base population trials in Argentina. Forest Genetics 9(3): 217-231.

Mann C.C., Plummer M.L. (2002) Forest biotech edges out of the lab. Science 295(5560): 1626-1629.

Marque C., El Kayal W., Debayles C., Keller G., Teulières C. (2003) Sucrose export defective gene isolated from *Eucalyptus*: A new gene involved in frost tolerance. *In* IUFRO Tree Biotechnology Meeting. Umeå, Sweden, Jun 7-12.

Marques C.M., Araujo J.A., Ferreira J.G., Whetten R., O'Malley D.M., Liu B.H., Sederoff R. (1998) AFLP genetic maps of *Eucalyptus globulus* and *E. tereticornis*. Theoretical and Applied Genetics 96(6-7): 727-737.

Marques C.M., Brondani R.P.V., Grattapaglia D., Sederoff R. (2002) Conservation and synteny of SSR loci and QTLs for vegetative propagation in four *Eucalyptus* species. Theoretical and Applied Genetics 105(2-3): 474-478.

Marques C.M., Vasquez-Kool J., Carocha V.J., Ferreira J.G., O'Malley D.M., Liu B.H., Sederoff R. (1999) Genetic dissection of vegetative propagation traits in *Eucalyptus tereticornis* and *E. globulus*. Theoretical and Applied Genetics 99(6): 936-946.

Marques C.M.P., Carocha V.J., Sousa A.M., Araujo J.A., Rodrigues J.C., Ribeiro M.M., Sanchez-Rodriguez L., Borallho N. (2003) *Eucalyptus* marker assisted breeding @ RAIZ. *In* IUFRO Tree Biotechnology Meeting. Umeå, Sweden, Jun 7-12.

Martens H., Naes T. (1991) Multivariate calibration. John Wiley and Sons, Chichester, UK.

Martin D., Rybicki E. (2000) RDP: detection of recombination amongst aligned sequences. Bioinformatics 16(6): 562-563.

Martin D.P., Posada D., Crandall K.A., Williamson C. (2005a) A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. AIDS Research and Human Retroviruses 21(1): 98-102.

Martin D.P., Williamson C., Posada D. (2005b) RDP2: Recombination detection and analysis from sequence alignments. Bioinformatics 21(2): 260-262.

Mason-Gamer R.J., Weil C.F., Kellogg E.A. (1998) Granule-bound starch synthase: Structure, function, and phylogenetic utility. Molecular Biology and Evolution 15(12): 1658-1673.

Mast A.R., Jones E.H., Havery S.P. (2005) An assessment of old and new DNA sequence evidence for the paraphyly of *Banksia* with respect to *Dryandra* (Proteaceae). Australian Systematic Botany 18: 75-88.

Maynard Smith J. (1992) Analyzing the mosaic structure of genes. Journal of Molecular Evolution 34: 126-129.

Maynard Smith J. (1999) The detection and measurement of recombination from sequence data. Genetics 153: 1021-1027.

McKie J.H., Jaouhari R., Douglas K.T., Goffner D., Feuillet C., Grima-Pettenati J., Boudet A.M., Baltas M., Gorrichon L. (1993) A molecular model for cinnamyl alcohol dehydrogenase, a plant aromatic alcohol dehydrogenase involved in lignification. Biochimica Et Biophysica Acta 1202(1): 61-69.

McKinnon G.E., Potts B.M., Steane D.A., Vaillancourt R.E. (in press) Population and phylogenetic analysis of the cinnamoyl coA reductase gene in *Eucalyptus globulus* (Myrtaceae). Australian Journal of Botany

McKinnon G.E., Steane D.A., Potts B.M., Vaillancourt R.E. (1999) Incongruence between chloroplast and species phylogenies in *Eucalyptus* subgenus *Monocalyptus* (Myrtaceae). American Journal of Botany 86(7): 1038-1046.

McKinnon G.E., Vaillancourt R.E., Jackson H.D., Potts B.M. (2001a) Chloroplast sharing in the Tasmanian eucalypts. Evolution 55(4): 703-711.

McKinnon G.E., Vaillancourt R.E., Steane D.A., Potts B.M. (2004) The rare silver gum, *Eucalyptus cordata*, is leaving its trace in the organellar gene pool of *Eucalyptus globulus*. Molecular Ecology 13: 3751-3762.

McKinnon G.E., Vaillancourt R.E., Tilyard P.A., Potts B.M. (2001b) Maternal inheritance of the chloroplast genome in *Eucalyptus globulus* and interspecific hybrids. Genome 44(5): 831-835.

Michell A.J. (1995) Pulpwood quality estimation by near-infrared spectroscopic measurements on eucalypt woods. Appita Journal 48: 425-428.

Milgate A.W., Vaillancourt R.E., Mohammed C., Powell M., Potts B.M. (in press) Genetic structure of a *Mycosphaerella cryptica* population Australasian Plant Pathology

Miller C.E. (1989) Analysis of synthetic polymers by near-infrared spectroscopy. PhD Thesis. University of Washington, Seattle, Washington.

Miranda I., Pereira H. (2001) Provenance effect on wood chemical composition and pulp yield for *Eucalyptus globulus* Labill. Appita Journal 54(4): 347-351.

Miranda I., Pereira H. (2002) Variation of pulpwood quality with provenances and site in *Eucalyptus globulus*. Annals of Forest Science 59(3): 283-291.

Missiaggia M., Piacezzi A., Grattapaglia D. (2002) A major effect QTL for early flowering in *Eucalyptus* mapped by selective genotyping of microsatellite markers detected in fluorescent multiplexes. *In* Proceedings of the 48th Brazilian Congress of Genetics. Aguas de Lindoir, Sept 17-20.

Mitchelson K.R., Chaparro J.X., Duong H., Yi J.-S., Jones M.E., Botero J., Teasdale R. (1997) Use of BSA for detailed genetic mapping in *Eucalyptus globulus*. *In* Plant & Animal Genome V Conference. San Diego, CA, Jan 12-16.

Mitchelson K.R., Drenth J., Duong H., Chaparro J.X. (1999) Direct sequencing of RAPD fragments using 3'-extended oligonucleotide primers and dye terminator cycle-sequencing. Nucleic Acids Research 27(19): e28.

Mohammed C., Barry K., Battaglia M., Beadle C., Eyles A., Mollon A., Pinkard E. (2000) Pruning-associated stem defects in plantation *E. nitens* and *E. globulus* grown for sawlog and veneer in Tasmania, Australia. *In* The Future of Eucalypts for Wood Production, pp 357-364. Launceston, Tasmania, Australia, Mar 19-24.

Moralejo M., Rochange F., Boudet A.M., Teulières C. (1998) Generation of transgenic *Eucalyptus globulus* plantlets through *Agrobacterium tumefaciens* mediated transformation. Australian Journal of Plant Physiology 25(2): 207-212.

Mori E.S., Zimback L., Moraes M.L.T., Rosa D.D., Velini E.D., Guerrini I.A., Marino C.L., Furtado E.L., Wilcken C.F., Maia I.G., Camargo L.E.A. (2004) Researching eucalypt expressed sequence tags of FORESTs database for plant growth hormone genes. *In* Plant & Animal Genomes XII Conference. San Diego, CA, Jan 10-14.

Mullins K.V., Llewellyn D.J., Hartney V.J., Strauss S., Dennis E.S. (1997) Regeneration and transformation of *Eucalyptus camaldulensis*. Plant Cell Reports 16(11): 787-791.

Muneri A., Raymond C.A. (2000) Genetic parameters and genotype-by-environment interactions for basic density, pilodyn penetration and stem diameter in *Eucalyptus globulus*. Forest Genetics 7(4): 317-328.

Muneri A., Raymond C.A. (2001) Nondestructive sampling of *Eucalyptus globulus* and *E. nitens* for wood properties; II. Fibre length and coarseness. Wood Science and Technology 35(1-2): 41-56.

Myburg A.A., Griffin A.R., Sederoff R.R., Whetten R.W. (2003) Comparative genetic linkage maps of *Eucalyptus grandis*, *Eucalyptus globulus* and their F1 hybrid based on a double pseudo-backcross mapping approach. Theoretical and Applied Genetics 107(6): 1028-1042.

Myburg A.A., Vogl C., Griffin A.R., Sederoff R.R., Whetten R.W. (2004) Genetics of postzygotic isolation in *Eucalyptus*: Whole-genome analysis of barriers to introgression in a wide interspecific cross of *Eucalyptus grandis* and *E. globulus*. Genetics 166(3): 1405-1418.

Nehls U., Beguiristain T., Ditengou F., Lapeyrie F., Martin F. (1998) The expression of a symbiosis-regulated gene in eucalypt roots is regulated by auxins and hypaphorine, the tryptophan betaine of the ectomycorrhizal basidiomycete *Pisolithus tinctorius*. Planta 207(2): 296-302.

Nelson N.D., Heather W.A. (1972) Wood color, basic density, and decay resistance in heartwood of fast-grown *Eucalyptus grandis* Hill ex. Maiden. Holzforschung 26: 54-60.

NIR Systems Inc. (1990) Manual for NIR spectral analysis software. Silver Spring, Maryland.

Nogueira L.M., Oda S., Mello E., Marino C.L. (2004) RAPD and SCAR markers used for tagging gene responsible for seedling anomaly in *Eucalyptus grandis*. *In* Proceedings of IUFRO Conference on *Eucalyptus* in a Changing World, edited by N.M.G. Borralho, J.S. Pereira, C. Marques, J. Coutinho, M. Madeira, M. Tome, pp 416-417. Aveiro, Portugal, Oct 11-15.

Norris K.H. (1989) Introduction and NIRS instrumentation. *In* Near infrared reflectance spectroscopy (NIRS): Analysis of forage quality, Vol 643. US Department of Agriculture, Agriculture Handbook, pp 6-21.

O'Connell A., Holt K., Piquemal J., Grima-Pettenati J., Boudet A., Pollet B., Lapierre C., Petit-Conil M., Schuch W., Halpin C. (2002) Improved paper pulp from plants with suppressed cinnamoyl-CoA reductase or cinnamyl alcohol dehydrogenase. Transgenic Research 11(5): 495-503.

Olsen K.M. (2002) Population history of *Manihot esculenta* (Euphorbiaceae) inferred from nuclear DNA sequences. Molecular Ecology 11: 901-911.

Olsen K.M., Schaal B.A. (1999) Evidence on the origin of cassava: Phylogeography of *Manihot esculenta*. Proceedings of the National Academy of Sciences USA 96: 5586-5591.

Okagaki R.J., Weil C.F. (1997) Analysis of recombination sites within the maize waxy locus. Genetics 147(2): 815-821.

Ona T., Sonoda T., Ito K., Shibata M. (1998) Relations between various extracted basic densities and wood chemical components in *Eucalyptus globulus*. Journal of Wood Science 44(3): 165-168.

Ona T., Sonoda T., Itoh K., Shibata M. (1997) Relationship of lignin content, lignin monomeric composition and hemicellulose composition in the same trunk sought by their within-tree variations in *Eucalyptus camaldulensis* and *E. globulus*. Holzforschung 51(5): 396-404.

Padidam M., Sawyer S., Fauquet C.M. (1999) Possible emergence of new geminiviruses by frequent recombination. Virology 265: 218-225.

Patterson G.I., Kubo K.M., Shroyer T., Chandler V.L. (1995) Sequences required for paramutation of the maize *b* gene map to a region containing the promoter and upstream sequences. Genetics 140: 1389-1406.

Paux E., Foucard C., Ladouce N., Sivadon P., Grima-Pettenati J., Marques C., Carocha V., Borallho N., Astorga R., Toval Hernandez G., Gion J.M., Vigneron P. (2003) Functional genomics of wood formation in *Eucalyptus*. *In* IUFRO Tree Biotechnology Meeting. Umeå, Sweden, Jun 7-12.

Paux E., Tamasloukht M., Ladouce N., Sivadon P., Grima-Pettenati J. (2004) Identification of genes preferentially expressed during wood formation in *Eucalyptus*. Plant Molecular Biology 55(2): 263-280.

Pearce R.B. (1996) Tansley review no. 87. Antimicrobial defences in the wood of living trees. New Phytologist 132(2): 203-233.

Pelosi A. (2002) Molecular and genetic studies into the formation of lateral roots in *Eucalyptus* and *Arabidopsis*. PhD Thesis. Monash University, Melbourne, Australia.

Pereira H., Sardinha R. (1984) Chemical composition of *Eucalyptus globulus* Lab. Appita Journal 37(8): 661-664.

Pfeil B.E., Crisp M.D. (2005) What to do with *Hibiscus*? A proposed nomenclatural resolution for a large and well known genus of Malvaceae and comments on paraphyly. Australian Systematic Botany 18: 49-60.

Pichon M., Courbou I., Beckert M., Boudet A.M., Grima-Pettenati J. (1998) Cloning and characterization of two maize cDNAs encoding Cinnamoyl-CoA Reductase (CCR) and differential expression of the corresponding genes. Plant Molecular Biology 38(4): 671-676.

Pinto G., Loureiro J., Lopes T., Santos C. (2004) Analysis of the genetic stability of *Eucalyptus globulus* Labill. somatic embryos by flow cytometry. Theoretical and Applied Genetics 109: 580-587.

Piquemal J., Lapierre C., Myton K., O'Connell A., Schuch W., Grima-Pettenati J., Boudet A.M. (1998) Down-regulation of cinnamoyl-CoA reductase induces significant changes of lignin profiles in transgenic tobacco plants. The Plant Journal 13(1): 71-83.

Plomion C., Leprovost G., Stokes A. (2001) Wood formation in trees. Plant Physiology 127: 1513-1523.

Plomion C., Pionneau C., Bailleres H. (2003) Analysis of protein expression along the normal to tension wood gradient in *Eucalyptus gunnii*. Holzforschung 57(4): 353-358.

Poeydomenge O., Boudet A.M., Grima-Pettenati J. (1994) A cDNA encoding S-adenosyl-L-methionine:caffeic acid 3-O-methyltransferase from *Eucalyptus*. Plant Physiology 105(2): 749-750.

Poeydomenge O., Marolda M., Boudet A.M., Grima-Pettenati J. (1995) Nucleotide sequence of a cDNA encoding mitochondrial malate dehydrogenase from *Eucalyptus*. Plant Physiology 107(4): 1455-1456.

Poke F.S. (2001) Lignin biosynthesis genes in *Eucalyptus*. Honours Thesis. University of Tasmania, Hobart, Tasmania.

Poke F.S., Vaillancourt R.E., Elliott R.C., Reid J.B. (2003) Sequence variation in two lignin biosynthesis genes, cinnamoyl CoA reductase (*CCR*) and cinnamyl alcohol dehydrogenase 2 (*CAD2*). Molecular Breeding 12(2): 107-118.

Poke F.S., Wright J.K., Raymond C.A. (2004) Predicting extractives and lignin contents in *Eucalyptus globulus* using near infrared reflectance analysis. Journal of Wood Chemistry and Technology 24(1): 55-67.

Posada D. (2002) Evaluation of methods for detecting recombination from DNA sequences: Empirical data. Molecular Biology and Evolution 19(5): 708-717.

Posada D., Crandall K.A. (2001) Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. Proceedings of the National Academy of Sciences USA 98(24): 13757-13762.

Posada D., Crandall K.A. (2002) The effect of recombination on the accuracy of phylogeny estimation. Journal of Molecular Evolution 54: 396-402.

Posada D., Crandall K.A., Holmes E.C. (2002) Recombination in evolutionary genomics. Annual Review of Genetics 36: 75-97.

Pot D., Chantre G., Rozenberg P., Rodrigues J.C., Jones G.L., Pereira H., Hannrup B., Cahalan C., Plomion C. (2002) Genetic control of pulp and timber properties in maritime pine (*Pinus pinaster* Ait.). Annals of Forest Science 59: 563-575.

Potts B.M. (2004) Genetic improvement of eucalypts. *In* Encyclopedia of Forest Sciences. Elsevier, Oxford, UK, pp 104-114.

Potts B.M., Barbour R.C., Hingston A., Vaillancourt R.E. (2003) Turner Review No. 6: Genetic pollution of native eucalypt gene pools - identifying the risks. Australian Journal of Botany 51: 1-25.

Potts B.M., Barbour R.C., Hingston A.B. (2001) Genetic pollution from farm forestry using eucalypt species and hybrids. Rural Industries Research and Development Corporation, Barton, ACT.

Potts B.M., Dungey H.S. (2004) Interspecific hybridization of eucalypts: key issues for breeders and geneticists. New Forests 27: 115-138.

Potts B.M., Dutkowski G.W., Jordan G.J., Vaillancourt R.E. (1999) Providing a population genetic framework for exploitation of eucalypt genetic resources: The case of *Eucalyptus globulus*. *In* Proceedings of the Australian Plant Breeding Conference, pp 97-101. Adelaide, South Australia, Apr 19-23.

Potts B.M., Wiltshire R.J.E. (1997) Eucalypt genetics and genecology. *In* J.E. Williams, J.C.Z. Woinarski, eds, Eucalypt ecology: Individuals to ecosystems. University Press, Cambridge, UK, pp 56-91.

Rafalski J.A. (2002) Novel genetic mapping tools in plants: SNPs and LD-based approaches. Plant Science 162(3): 329-333.

Ranik M., Creux N., Myburg A.A. (2004) Transcriptome analysis of xylogenesis in *Eucalyptus* using cDNA-AFLP and Li-Cor automated DNA analysers. *In* Plant & Animal Genomes XII Conference. San Diego, CA, Jan 10-14.

Rathbone D.A. (2003) Genetic differentiation in a regionally rare eucalypt, *Eucalyptus perriniana*. Honours Thesis. University of Tasmania, Hobart, Australia.

Raymond C.A., Banham P., MacDonald A.C. (1998) Within tree variation and gentic control of basic density, fibre length and coarseness in *Eucalyptus regnans* in Tasmania. Appita Journal 51(4): 299-305.

Raymond C.A., MacDonald A.C. (1998) Where to shoot your pilodyn: within tree variation in basic density in plantation *Eucalyptus globulus* and *E. nitens* in Tasmania. New Forests 15: 205-221.

Raymond C.A., Muneri A. (2001) Nondestructive sampling of *Eucalyptus globulus* and *E. nitens* for wood properties. I. Basic density. Wood Science and Technology 35(1-2): 27-39.

Raymond C.A., Schimleck L.R. (2002) Development of near infrared reflectance analysis calibrations for estimating genetic parameters for cellulose content in *Eucalyptus globulus*. Canadian Journal of Forest Research 32(1): 170-176.

Raymond C.A., Schimleck L.R., Muneri A., Michell A.J. (2001a) Genetic parameters and genotype-by-environment interactions for pulp yield predicted using near infrared reflectance analysis and pulp productivity in *Eucalyptus globulus*. Forest Genetics 8(3): 213-224.

Raymond C.A., Schimleck L.R., Muneri A., Michell A.J. (2001b) Nondestructive sampling of *Eucalyptus globulus* and *E. nitens* for wood properties. III. Predicted pulp yield using near infrared reflectance analysis. Wood Science and Technology 35(3): 203-215.

Rodrigues J., Graca J., Pereira H. (2001) Influence of tree eccentric growth on syringyl/guaiacyl ratio in *Eucalyptus globulus* wood lignin assessed by analytical pyrolysis. Journal of Analytical and Applied Pyrolysis 58: 481-489.

Rodrigues J., Meier D., Faix O., Pereira H. (1999) Determination of tree to tree variation in syringyl/guaiacyl ratio of *Eucalyptus globulus* wood lignin by analytical pyrolysis. Journal of Analytical and Applied Pyrolysis 48(2): 121-128.

Saka S. (1984) The relationship between the microfibril orientation in the tracheid S2 layer and lignin content of coniferous woods. *In* Proceedings of the Pacific Regional Wood Anatomy Conference, edited by S. Sudo, pp 100-102. Tsukuba, Ibaraki, Japan, Oct 1-7.

Sale M.M., Potts B.M., West A.K., Reid J.B. (1993) Relationships within *Eucalyptus* using chloroplast DNA. Australian Systematic Botany 6: 127-138.

Sale M.M., Potts B.M., West A.K., Reid J.B. (1996) Relationships within *Eucalyptus* (Myrtaceae) using PCR-amplification and southern hybridisation of chloroplast DNA. Australian Systematic Botany 9(3): 273-282.

Sang T. (2002) Utility of low-copy nuclear gene sequences in plant phylogenetics. Critical Reviews in Biochemistry and Molecular Biology 37(3): 121-147.

Sato S., Horikiri K., Miyashita K., Ishige N., Hibino T., Kawazu T. (2001) Analysis of wood development with a genomic approach: *Eucalyptus* ESTs and TAC genomic library. *In* Plant & Animal Genome IX Conference. San Diego, CA, Jan 13-17.

Sawbridge T.I., Drenth J., Anderson M., Hallinan J., Norris B., Teasdale R. (1999) Applications of EST sequencing in a forestry biotech company. *In* Plant & Animal Genome VII Conference. San Diego, CA, Jan 17-21.

Schimleck L.R., Evans R., Ilic J. (2001) Estimation of *Eucalyptus delegatensis* wood properties by near infrared spectroscopy. Canadian Journal of Forest Research 31: 1671-1675.

Schimleck L.R., Evans R., Matheson A.C. (2002) Estimation of *Pinus radiata* D. Don clear wood properties by near-infrared spectroscopy. Journal of Wood Science 48(2): 132-137.

Schimleck L.R., Michell A.J. (1998) Determination of within-tree variation of kraft pulp yield using near-infrared spectroscopy. TAPPI Journal 81: 229-236.

Schimleck L.R., Michell A.J., Raymond C.A., Muneri A. (1999) Rapid assessment of pulpwood quality using near-infrared spectroscopy. *In* 9th International Conference on Near-infrared Spectroscopy, edited by A.M.C. Daries, R. Giangiacomo. Verona, Italy, Jun 14-18.

Shani Z., Dekel M., Cohen B., Barimboim N., Kolosovski N., Safranuvitch A., Cohen O., Shoseyov O. (2003) Cell wall modification for the enhancement of commercial *Eucalyptus* species. *In* IUFRO Tree Biotechnology Meeting. Umeå, Sweden, Jun 7-12.

Shao Z., Chen W., Luo H., Ye X., Zhang J. (2002) Studies on the introduction of the cecropin D gene into *Eucalyptus urophylla* to breed the resistant

varieties to *Pseudomonas solanacearum*. Scientia Silvae Sinicae 38(2): 92-97.

Shepherd M., Chaparro J., Dale G., Jefferson L., Duong H., Vogel H., Walsh J., Gibbings M., Teasdale R. (1995) Mapping insect resistance and essential oil traits in a tropical *Eucalyptus* hybrid. *In* CRCTHF IUFRO Conference on Eucalypt Plantations: Improving Fibre Yield and Quality, pp 420-423. Hobart, Tasmania, Australia, Feb 19-24.

Shepherd M., Chaparro J.X., Teasdale R. (1999) Genetic mapping of monoterpene composition in an interspecific eucalypt hybrid. Theoretical and Applied Genetics 99(7-8): 1207-1215.

Shepherd M., Jones M.E. (2005) Molecular markers in tree improvement: Characterisation and use in *Eucalyptus*. *In* H. Lorz, G. Wenzel, eds, Molecular marker systems in plant breeding and crop improvement. Springer-Verlag, Heidelberg, Germany, pp 399-409.

Small R.L., Cronn R.C., Wendel J.F. (2004) Use of nuclear genes for phylogeny reconstruction in plants. Australian Systematic Botany 17(2): 145-170.

Small R.L., Ryburn J.A., Cronn R.C., Seelanan T., Wendel J.F. (1998) The tortoise and the hare: choosing between noncoding plastome and nuclear *Adh* sequences for phylogeny reconstruction in a recently diverged plant group. American Journal of Botany 85: 1301-1315.

Smook G. (1992) Handbook for pulp & paper technologists, Ed Second. Angus Wilde Publications, Vancouver, B.C.

Smouse P.E. (2000) Reticulation inside the species boundary. Journal of Classification 17(2): 165-173.

Song Y., Cullis C.A. (1992) RFLP and RAPD mapping of *Eucalyptus globulus*. *In* Plant Genome I Conference. San Diego, CA, Nov.

Sosef M.S.M. (1997) Hierarchical models, reticulate evolution and the inevitability of paraphyletic taxa. Taxon 46: 75-85.

Southam C.M., Ehrlich J. (1943) Decay resistance and physical characteristics of wood. Journal of Forestry 41: 666-673.

Southerton S.G., Marshall H., Mouradov A., Teasdale R.D. (1998a) Eucalypt MADS-box genes expressed in developing flowers. Plant Physiology 118(2): 365-372.

Southerton S.G., Strauss S.H., Olive M.R., Dennis E.S. (1993) Cloning of the floral homeotic gene homolog of *Leafy Floricaula* in *Eucalyptus globulus*. Journal of Cellular Biochemistry Suppl. 17B: 24.

Southerton S.G., Strauss S.H., Olive M.R., Harcourt R.L., Decroocq V., Zhu X.M., Llewellyn D.J., Peacock W.J., Dennis E.S. (1998b) *Eucalyptus* has a functional equivalent of the *Arabidopsis* floral meristem identity gene *LEAFY*. Plant Molecular Biology 37(6): 897-910.

Sreenivasulu N., Kishor P.B.K., Varshney R.K., Altschmied L. (2002) Mining functional information from cereal genomes - the utility of expressed sequence tags. Current Science 83(8): 965-973.

Steane D.A. (2005) Complete nucleotide sequence of the chloroplast genome from the Tasmanian Blue Gum, *Eucalyptus globulus* (Myrtaceae). DNA Research 12(3): 215-220.

Steane D.A., Byrne M., Vaillancourt R.E., Potts B.M. (1998) Chloroplast DNA polymorphism signals complex interspecific interactions in *Eucalyptus* (Myrtaceae). Australian Systematic Botany 11(1): 25-40.

Steane D.A., McKinnon G.E., Vaillancourt R.E., Potts B.M. (1999) ITS sequence data resolve higher level relationships among the eucalypts. Molecular Phylogenetics and Evolution 12(2): 215-223.

Steane D.A., Nicolle D., McKinnon G.E., Vaillancourt R.E., Potts B.M. (2002) Higher-level relationships among the eucalypts are resolved by ITS-sequence data. Australian Systematic Botany 15(1): 49-62.

Steane D.A., Vaillancourt R.E., Russell J., Powell W., Marshall D., Potts B.M. (2001) Development and characterisation of microsatellite loci in *Eucalyptus globulus* (Myrtaceae). Silvae Genetica 50(2): 89-91.

Steane D.A., West A.K., Potts B.M., Ovenden J.R., Reid J.B. (1991) Restriction fragment length polymorphisms in chloroplast DNA from 6 species of *Eucalyptus*. Australian Journal of Botany 39(5): 399-414.

Stirling B., Yang Z.K., Gunter L.E., Tuskan G.A., Bradshaw H.D. (2003) Comparative sequence analysis between orthologous regions of the *Arabidopsis* and *Populus* genomes reveals substantial synteny and microcollinearity. Canadian Journal of Forest Research 33(11): 2245-2251.

Stokoe R.L., Shepherd M., Lee D.J., Nikles D.G., Henry R.J. (2001) Natural inter-subgeneric hybridization between *Eucalyptus acmenoides* Schauer and

Eucalyptus cloeziana F. Muell (*Myrtaceae*) in southeast Queensland. Annals of Botany 88(4): 563-570.

Strabala T. (2004) Expressed sequence tag databases from forestry tree species. *In* S. Kumar, M. Fladung, eds, in Molecular genetics and breeding of forest trees. The Haworth Press, Inc., Binghamton, NY

Swofford D.L. (1999) PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods), Version 4. Sinauer, Sunderland, MA.

Tagu D., Martin F. (1995) Expressed sequence tags of randomly selected cDNA clones from *Eucalyptus globulus-Pisolithus tinctorius* ectomycorrhiza. Molecular Plant-Microbe Interactions 8(5): 781-783.

Tank D.C., Sang T. (2001) Phylogenetic utility of the glycerol-3-phosphate acyltransferase gene: Evolution and implications in *Paeonia* (Paeoniaceae). Molecular Phylogenetics and Evolution 19(3): 421-429.

TAPPI (1989) Basic density and moisture content of pulpwood. TAPPI no. T258 om-98

Tenaillon M.I., Sawkins M.C., Long A.D., Gaut R.L., Doebley J.F., Gaut B.S. (2001) Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). Proceedings of the National Academy of Sciences USA 98(16): 9161-9166.

Terdwongworakul A., Punsuwan V., Thanapase W., Tsuchikawa S. (2005) Rapid assessment of wood chemical properties and pulp yield of *Eucalyptus camaldulensis* in Thailand tree plantations by near infrared spectroscopy for improving wood selection for high quality pulp. Journal of Wood Science 51(2): 167-171.

Thamarus K., Groom K., Bradley A., Raymond C.A., Schimleck L.R., Williams E.R., Moran G.F. (2004) Identification of quantitative trait loci for wood and fibre properties in two full-sib pedigrees of *Eucalyptus globulus*. Theoretical and Applied Genetics 109(4): 856-864.

Thamarus K.A., Groom K., Murrell J., Byrne M., Moran G.F. (2002) A genetic linkage map for *Eucalyptus globulus* with candidate loci for wood, fibre, and floral traits. Theoretical and Applied Genetics 104(2-3): 379-387.

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408: 796-815.

Thomas M.A., Klaper R. (2004) Genomics for the ecological toolbox. Trends in Ecology and Evolution 19(8): 439-445.

Thumma B.R., Nolan M.F., Evans R., Moran G.F. (2005) Polymorphisms in *cinnamoyl CoA reductase* (*CCR*) are associated with variation in microfibril angle in *Eucalyptus* spp. Genetics 171: 1257-1265.

Tournier V., Grat S., Marque C., El Kayal W., Penchel R., de Andrade G., Boudet A.M., Teulières C. (2003) An efficient procedure to stably introduce genes into an economically important pulp tree (*Eucalyptus grandis* x *Eucalyptus urophylla*). Transgenic Research 12(4): 403-411.

Trigueiro R.M., Guerrini I.A., Leite R.M., Maia I.G., Marino C.L., Mori E.S., Velini E.D., Camargo L.E.A., Furtado E.L., Wilcken C.F. (2004) Identification of eucalypt ESTs involved in proline production. *In* Plant & Animal Genomes XII Conference. San Diego, CA, Jan 10-14.

Turnbull J.W. (1999) Eucalypt plantations. New Forests 17(1-3): 37-52.

Tuskan G.A., DiFazio S.P., Teichmann T. (2004) Poplar genomics is getting popular: The impact of the poplar genome project on tree research. Plant Biology 6(1): 2-4.

Udovicic F., Ladiges P.Y. (2000) Informativeness of nuclear and chloroplast DNA regions and the phylogeny of the eucalypts and related genera (Myrtaceae). Kew Bulletin 55: 633-645.

Udovicic F., McFadden G.I., Ladiges P.Y. (1995) Phylogeny of *Eucalyptus* and *Angophora* based on 5S rDNA spacer sequence data. Molecular Phylogenetics and Evolution 4(3): 247-256.

Vaillancourt R.E., Jackson H.D. (2000) A chloroplast DNA hypervariable region in eucalypts. Theoretical and Applied Genetics 101(3): 473-477.

Vaillancourt R.E., Petty A., McKinnon G.E. (2004) Maternal inheritance of mitochondria in *Eucalyptus globulus*. Journal of Heredity 95: 353-355.

Vaillancourt R.E., Potts B.M., Manson A., Reid J.B. (1994) Detection of QTLs in a *Eucalyptus gunnii* x *E. globulus* F2 using a RAPD linkage map. *In* Proceedings of the International Wood Biotechnology Symposium, edited by N. Morohoshi, pp 63-70. Tokyo, Japan,

Valerio L., Carter D., Rodrigues J.C., Tournier V., Gominho J., Marque C., Boudet A.M., Maunders M., Pereira H., Teulières C. (2003) Down regulation of

cinnamyl alcohol dehydrogenase, a lignification enzyme, in *Eucalyptus camaldulensis*. Molecular Breeding 12(2): 157-167.

Van Der Nest M.A., Steenkamp E.T., Wingfield B.D., Wingfield M.J. (2000) Development of simple sequence repeat (SSR) markers in *Eucalyptus* from amplified inter-simple sequence repeats (ISSR). Plant Breeding 119(5): 433-436.

Velini E.D., Trindade M.L.B., Alves E., Marino C.L., Mori E.S., Guerrini I.A., Furtado E.L., Wilcken C.F., Maia I.G. (2004) ESTs of *Eucalyptus* related with resistance of herbicide inhibitors of ALS (acetolactate synthase) and EPSPS (5-enolpyrovylshikimic acid-3-phosphate) enzymes. *In* Plant & Animal Genomes XII Conference. San Diego, CA, Jan 10-14.

Verhaegen D., Plomion C. (1996) Genetic mapping in *Eucalyptus urophylla* and *Eucalyptus grandis* using RAPD markers. Genome 39(6): 1051-1061.

Verhaegen D., Plomion C., Gion J.M., Poitel M., Costa P., Kremer A. (1997) Quantitative trait dissection analysis in *Eucalyptus* using RAPD markers.1. Detection of QTL in interspecific hybrid progeny, stability of QTL expression across different ages. Theoretical and Applied Genetics 95(4): 597-608.

Voiblet C., Duplessis S., Encelot N., Martin F. (2001) Identification of symbiosis-regulated genes in *Eucalyptus globulus-Pisolithus tinctorius* ectomycorrhiza by differential hybridization of arrayed cDNAs. The Plant Journal 25(2): 181-191.

Walker J.C.F., Butterfield B.G. (1995) The importance of microfibril angle for the processing industries. New Zealand Forestry 40(4): 34-40.

Wallis A.F.A., Wearne R.H., Wright P.J. (1996) Analytical characteristics of plantation eucalypt woods relating to kraft pulp yields. Appita Journal 49(6): 427-432.

Wallis A.F.A., Wearne R.H., Wright P.J. (1997) New approaches to rapid analysis of cellulose in wood. *In* Proceedings of the International Symposium on Wood and Pulping Chemistry, pp 1-4. Montreal, Quebec, June.

Wang X., Hughes A.L., Tsukamoto T., Ando T., Kao T. (2001) Evidence that intragenic recombination contributes to allelic diversity of the S-RNase gene at the self-incompatibility (S) locus in *Petunia inflata*. Plant Physiology 125(2): 1012-1022.

Wang X.Q., Tank D.C., Sang T. (2000) Phylogeny and divergence time in Pinaceae: evidence from three genomes. Molecular Biology and Evolution 17: 773-781.

Wardlaw T., Mohammed C., Barry K., Eyles A., Wiseman D., Beadle C., Battaglia M., Pinkard L., Kube P. (2003) Interdisciplinary approach to the study and management of stem defect in eucalypts. New Zealand Journal of Forestry Science 33(3): 385-398.

Washusen R., Ades P., Evans R., Ilic J., Vinden P. (2001) Relationships between density, shrinkage, extractives content and microfibril angle in tension wood from three provenances of 10-year-old *Eucalyptus globulus* Labill. Holzforschung 55(2): 176-182.

Watson J.M., Brill E.M. (2004) *Eucalyptus grandis* has at least two functional *SOC1*-like floral activator genes. Functional Plant Biology 31(3): 225-234.

White D.A., Raymond C.A., Kile G.A., Hall M.F. (1999) Are there genetic differences in susceptibility of *Eucalyptus nitens* and *E. regnans* stems to defect and decay? Australian Forestry 62(4): 368-374.

Whittock S., Steane D.A., Vaillancourt R.E., Potts B.M. (2003) Molecular evidence shows that the tropical boxes (*Eucalyptus* subgenus *Minutifructus*) are over-ranked. Transactions of the Royal Society of South Australia 127: 27-32.

Wilcken C.F., Firmino D.C., Rosa D.D., Correa E.A., Furtado E.L., Velini E.D., Mori E.S., Guerrini I.A., Marino C.L., Camargo L.E.A. (2004) DOXP/MEP pathway enzymes genes annotated in *Eucalyptus* expressed sequence tags of FORESTs database. *In* Plant & Animal Genomes XII Conference. San Diego, CA, Jan 10-14.

Williams J.E., Woinarski J.C.Z. (1997) Eucalypt Ecology: Individuals to ecosystems. Cambridge University Press, Cambridge, UK.

Wiuf C., Christensen T., Hein J. (2001) A simulation study of the reliability of recombination detection methods. Molecular Biology and Evolution 18(10): 1929-1939.

Wong A.H.H., Wilkes J., Heather W.A. (1983) Influence of wood density and extractives content on the decay resistance of the heartwood of *Eucalyptus delegatensis* R. T. Baker. Journal of the Institute of Wood Science 9(6): 261-263.

Workman J.J.J. (1992) NIR spectroscopy calibration basics. *In* D.A. Burns, E.W. Ciurczak, eds, Handbook of near-infrared analysis. Marcel Dekker, Inc., New York, pp 274-276.

Wright J.A., Birkett M.D., Gambino M.J.T. (1990) Prediction of pulp yield and cellulose content from wood samples using near infrared reflectance spectroscopy. TAPPI Journal 73: 164-166.

Xu X., Hsia A.-P., Zhang L., Nikolau B.J., Schnable P.S. (1995) Meiotic recombination break points resolve at high rates at the 5' end of a maize coding sequence. Plant Cell 7: 2151-2161.

Yamada-Watanabe K., Kawaoka A., Matsunaga E., Nanto K., Sugita K., Endo S., Ebinuma H., Murata N. (2003) Molecular breeding of *Eucalyptus* - Analysis of salt stress tolerance of transgenic *Eucalyptus camaldulensis* that overexpress choline oxidase gene (*cod A*). *In* IUFRO Tree Biotechnology Meeting. Umeå, Sweden, Jun 7-12.

Yeh T.F., Chang H.M., Kadla J.F. (2004) Rapid prediction of solid wood lignin content using transmittance near-infrared spectroscopy. Journal of Agricultural and Food Chemistry 52: 1435-1439.

Yokoi H., Ishida Y., Ohtani H., Tsuge S., Sonoda T., Ona T. (1999) Characterization of within-tree variation of lignin components in *Eucalyptus camaldulensis* by pyrolysis-gas chromatography. Analyst 124(5): 669-674.

Yu J., Hu S.N., Wang J., Wong G.K.S., Li S.G., Liu B., Deng Y.J., Dai L., Zhou Y., Zhang X.Q., Cao M.L., Liu J., Sun J.D., Tang J.B., Chen Y.J., Huang X.B., Lin W., Ye C., Tong W., Cong L.J., Geng J.N., Han Y.J., Li L., Li W., Hu G.Q., Huang X.G., Li W.J., Li J., Liu Z.W., Liu J.P., Qi Q.H., Liu J.S., Li T., Wang X.G., Lu H., Wu T.T., Zhu M., Ni P.X., Han H., Dong W., Ren X.Y., Feng X.L., Cui P., Li X.R., Wang H., Xu X., Zhai W.X., Xu Z., Zhang J.S., He S.J., Zhang J.G., Xu J.C., Zhang K.L., Zheng X.W., Dong J.H., Zeng W.Y., Tao L., Ye J., Tan J., Ren X.D., Chen X.W., He J., Liu D.F., Tian W., Tian C.G., Xia H.G., Bao Q.Y., Li G., Gao H., Cao T., Zhao W.M., Li P., Chen W., Wang X.D., Zhang Y., Hu J.F., Liu S., Yang J., Zhang G.Y., Xiong Y.Q., Li Z.J., Mao L., Zhou C.S., Zhu Z., Chen R.S., Hao B.L., Zheng W.M., Chen S.Y., Guo W., Li G.J., Liu S.Q., Tao M., Zhu L.H., Yuan L.P., Yang H.M. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp *indica*). Science 296(5565): 79-92.

Zhang D., Hewitt G.M. (2003) Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. Molecular Ecology 12: 563-584.

Zik M., Irish V.F. (2003) Flower development: Initiation, differentiation, and diversification. Annual Review of Cell and Developmental Biology 19: 119-140.

Zimback L., Mori E.S., Moraes M.L.T., Rosa D.D., Furtado E.L., Marino C.L., Wilcken C.F., Velini E.D., Guerrini I.A., Maia I.G., Camargo L.E.A. (2004) Data mining of *Eucalyptus* ESTs involved in the mechanisms non hormonal growth genes. *In* Plant & Animal Genomes XII Conference. San Diego, CA, Jan 10-14.

APPENDIX 1

Supporting material written during candidature

Poke F.S., Vaillancourt R.E., Elliott R.C., Reid J.B. (2003)
Sequence variation in two lignin biosynthesis genes, cinnamoyl CoA reductase (*CCR*) and cinnamyl alcohol dehydrogenase 2 (*CAD2*).
Molecular Breeding 12(2): 107-118.

Chapter has been removed for copyright or proprietary reasons.